DOCUMENT RESUME

ED 318 057 CS 507 128

AUTHOR	Pisoni, David B.; And Others
TITLE	Research on Speech Perception. Progress Report No. 13.
INSTITUTION	Indiana Univ., Bloomington. Dept. of Psychology.
SPONS AGENCY	National Institutes of Health (DHHS), Bethesda, Md.; National Science Foundation, Washington, D.C.
PUB DATE	87
CONTRACT	AF-F-33615-86-K-0549
GRANT	IRI-86-17847; NS-07134-09; NS-12179-11
NOTE	337p.; For other reports in this series, see CS 507 123-129.
PUB TYPE	Reports - Research/Technical (143) -- Collected Works - General (020) -- Information Analyses (070)
EDRS PRICE	MF01/PC14 Plus Postage.
DESCRIPTORS	*Acoustic Phonetics; Auditory Discrimination; *Auditory Perception; Communication Research; Computer Software Development; Infants; *Language Processing; Language Research; Linguistics; Speech; Speech Synthesizers
IDENTIFIERS	Indiana University Bloomington; *Speech Perception; Speech Research; Theory Development

ABSTRACT

Summarizing research activities in 1987, this is the thirteenth annual report of research on speech perception, analysis, synthesis, and recognition conducted in the Speech Research Laboratory of the Department of Psychology at Indiana University. The report includes extended manuscripts, short reports, progress reports, and information on instrumentation developments and software support. The report contains the following 15 articles: "Some Effects of Talker Variability on Spoken Word Recognition" (J. W. Mullenix and others); "Effects of Talker Variability on Recall of Spoken Word Lists" (C. S. Martin and others); "The Perception of Digitally Coded Speech by Native and Non-Native Speakers of English" (K. Ozawa and J. S. Logan); "F1 Structure Provides Information for Final-Consonant Voicing" (W. V. Summers); "Comparative Research on Language Learning" (J. A. Gierut); "Maximal Opposition Approach to Phonological Treatment" (J. A. Gierut); "The Effects of Semantic Context on Voicing Neutralization" (J. Charles-Luce); "Stimulus Variability and Processing Dependencies in Speech Perception" (J. W. Mullenix and D. B. Pisoni); "Some Observations concerning English Stress and Phonotactics Using a Computerized Lexicon" (S. Davis); "External Validity of Productive Phonological Knowledge: A First Report" (J. A. Gierut and others); "Effects of Changes in Spectral Slope on the Intelligibility of Speech in Noise" (R. I. Pedlow); "On the Arguments for Syllable-Internal Structure" (S. Davis); "The Identification of Speech Using Word and Phoneme Labels" (J. S. Logan); "Talker Variability and the Recall of Spoken Word Lists: A Replication and Extension" (J. S. Logan and D. B. Pisoni); and "SAP: A Speech Acquisition Program for the SRL-VAX" (M. J. Dedina). Lists of publications and of laboratory staff, associated faculty and personnel conclude the report. (SR)

ED318057

RESEARCH ON SPEECH PERCEPTION

Progress Report No. 13
(1987)

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana
47405

Supported by

Department of Health and Human Services
U.S. Public Health Service

National Institutes of Health
Research Grant No. NS-12179-11

National Institutes of Health
Training Grant No. NS-07134-09

National Science Foundation
Research Grant No. IRI-86-17847

and

U.S. Air Force
Armstrong Aerospace Medical Research Laboratory
Contract No. AF-F-33615-86-C-0549

CS507128

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

BEST COPY AVAILABLE

RESEARCH ON SPEECH PERCEPTION

Progress Report No. 13
(1987)

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana
47405

Supported by

Department of Health and Human Services
U.S. Public Health Service

National Institutes of Health
Research Grant No. NS-12179-11

National Institutes of Health
Training Grant No. NS-07134-09

National Science Foundation
Research Grant No. IRI-86-17847

and

U.S. Air Force
Armstrong Aerospace Medical Research Laboratory
Contract No. AF-F-33615-86-C-0549

RESEARCH ON SPEECH PERCEPTION

Progress Report No. 13
(1987)

David B. Pisoni, Ph.D.
Principal Investigator

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

Research Supported by:

Department of Health and Human Services
U. S. Public Health Service

National Institutes of Health
Research Grant No. NS-12179-11

National Institutes of Health
Training Grant No. NS-07134-09

National Science Foundation
Research Grant No. IRI-86-17847

and

U. S. Air Force
Armstrong Aerospace Medical Research Laboratory
Contract No. AF-F-33615-86-C-0549

[RESEARCH ON SPEECH PERCEPTION Progress Report No. 13 (1987)]

Table of Contents

Introduction iii

I. Extended Manuscripts 1

Some effects of talker variability on spoken word recognition; John W. Mullennix, David B. Pisoni, and Christopher S. Martin . . 3

Effects of talker variability on recall of spoken word lists; Christopher S. Martin, John W. Mullennix, David B. Pisoni, and W. Van Summers 41

The perception of digitally coded speech by native and non-native speakers of English; Kazunori Ozawa and John S. Logan 71

F1 structure provides information for final-consonant voicing; W. Van Summers 101

Comparative research on language learning; Judith A. Gierut 121

Maximal opposition approach to phonological treatment; Judith A. Gierut 139

The effects of semantic context on voicing neutralization; Jan Charles-Luce 167

Stimulus variability and processing dependencies in speech perception; John W. Mullennix and David B. Pisoni; 197

II. Short Reports and Work in Progress 223

Some observations concerning English stress and phonotactics using a computerized lexicon; Stuart Davis 225

External validity of productive phonological knowledge: A first report; Judith A. Gierut, Daniel A. Dinnsen, and Kathleen Bardovi-Harlig; 241

Effects of changes in spectral slope on the intelligibility of speech in noise; Robert I. Pedlow 249

On the arguments for syllable-internal structure; Stuart Davis 265

The identification of speech using word and phoneme labels; John S. Logan. 277

Talker variability and the recall of spoken word lists: A replication and extension; John S. Logan and David B. Pisoni. 307

III. Instrumentation and Software Development 329

SAP: A speech acquisition program for the SRL-VAX; Michael J. Dedina. 331

IV. Publications 339

V. SRL Laboratory Staff and Personnel 343

INTRODUCTION

This is the thirteenth annual report summarizing the research activities on speech perception, analysis, synthesis, and recognition carried out in the Speech Research Laboratory, Department of Psychology, Indiana University in Bloomington. As with previous reports, our main goal has been to summarize various research activities over the past year and make them readily available to granting agencies, sponsors and interested colleagues in the field. Some of the papers contained in this report are extended manuscripts that have been prepared for formal publication as journal articles or book chapters. Other papers are simply short reports of research presented at professional meetings during the past year or brief summaries of "on-going" research projects in the laboratory. From time to time, we also have included new information on instrumentation and software support when we think this information would be of interest or help to others. We have found the sharing of this information to be very useful in facilitating our own research.

We are distributing reports of our research activities because of the ever increasing lag in journal publications and the resulting delay in the dissemination of new information and research findings in the field of speech processing. We are, of course, very interested in following the work of other colleagues who are carrying out research on speech perception, production, analysis, synthesis, and recognition and, therefore, we would be grateful if you would send us copies of your own recent reprints, preprints and progress reports as they become available so that we can keep up with your latest findings. Please address all correspondence to:

Professor David B. Pisoni
Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405
USA
(812) 335-1155

Copies of this report are being sent primarily to libraries and specific research institutions rather than individual scientists. Because of the rising costs of publication and printing, it is not possible to provide multiple copies of this report to people at the same institution or issue copies to individuals. We are eager to enter into exchange agreements with other institutions for their reports and publications. Please write to the above address.

The information contained in the report is freely available to the public and is not restricted in any way. The views expressed in these research reports are those of the individual authors and do not reflect the opinions of the granting agencies or sponsors of the specific research.

I. EXTENDED MANUSCRIPTS

[RESEARCH ON SPEECH PERCEPTION Progress Report No. 13 (1987) Indiana University]

Some Effects of Talker Variability on Spoken Word Recognition*

John W. Mullennix, David B. Pisoni, and Christopher S. Martin

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, IN 47405

---

*The research reported here was supported by NIH Research Grant NS-12179-11 and NIH Training Grant NS-07134-09 to Indiana University in Bloomington. The authors would like to thank Paul A. Luce and W. Van Summers for helpful suggestions, comments, and criticisms. An earlier version of Experiment 1 was previously reported in Progress Report No. 12.

Abstract

The perceptual consequences of trial-to-trial changes in the voice of the talker on spoken word recognition were examined. The results from a series of experiments using perceptual identification and naming tasks demonstrated that perceptual performance decreases when the voice of the talker changed from trial-to-trial compared to performance when the voice on each trial remains the same. In addition, the effects of talker variability on word recognition appeared to be more robust and less dependent on the type of task than the effects of word frequency and lexical structure. Possible hypotheses regarding the nature of the processes giving rise to these effects are discussed, with particular attention to the idea that the processing of information about the talker's voice is intimately related to early perceptual processes that extract acoustic-phonetic information from the speech signal.

Some Effects of Talker Variability on Spoken Word Recognition

One of the most important unresolved problems in human speech perception concerns perceptual normalization. The speech signal is characterized by extreme variability in its acoustic composition. The acoustic cues to consonants and vowels vary as a function of phonetic, phonological, lexical, and sentential context, speaking rate, individual talker characteristics, and many other factors. Although the acoustic parameters specifying a particular utterance vary as a function of these factors, utterances elicited under a variety of conditions and from a variety of speakers are readily perceived and understood quite easily by the average listener without conscious awareness of the source characteristics. This general observation has led researchers and theorists to assume that a perceptual process or mechanism may exist to automatically "adjust" or "normalize" the acoustic differences between utterances in order to preserve perceptual constancy of the linguistic message. At the present time, very little is known about the nature of perceptual normalization in speech. Furthermore, the perceptual consequences of normalization and its impact on other cognitive processes involved in the perception of spoken language have not been extensively studied either. In the present investigation, we focused on one particular factor involved in normalization, namely, the effects of talker variability on perception. We hoped that by studying the effects of changes in a talker's voice from trial to trial we would be able to learn more about the underlying normalization processes in speech perception (see reference note 1).

Differences in voice characteristics among individual talkers may be due to a wide variety of factors. Structural factors related to the physical shape and length of the oral and nasal vocal tract cavities constrain the ultimate acoustic composition of the speech signal. This may be illustrated by considering the differences in vocal tract size, length and shape between men, women, and children and how these differences affect the formant frequencies of vowels (Peterson & Barney, 1952). These structural differences result in large variations in voice characteristics between talkers. One consequence of this is that the acoustic properties of vowels produced by different talkers may vary substantially (e.g. see Fant, 1973; Joos, 1948; Peterson & Barney, 1952). Differences in the glottal source function also exist between talkers, resulting in other voice quality differences that distinguish speakers (see Carr & Trill, 1964; Carrell, 1984; Monsen & Engebretson, 1977). In addition to anatomical or structural factors, a number of more dynamic factors also affect the speech signal, such as the control and positioning of the articulators and the manner in which the vocal gestures are carried out (Ladefoged, 1980). Individual talkers produce vowels and consonants differently, as reflected by differences between talkers in acoustic measures such as short-term energy spectra, fundamental frequency contours, durations, and the length and rate of change of formant transitions.

Given the substantial acoustic differences between talkers, the problem of compensating for these sources of variability in perception becomes an important and fundamental research issue. Joos (1948) was among the very first researchers to address this issue in any detail in his classic monograph. He proposed that the perception of vowels not only depends on the absolute formant frequencies of the vowel but also on the relationship of these values to those of the formant frequencies for other vowels produced by the same talker. Ladefoged and Broadbent (1957) provided perceptual evidence supporting this hypothesis. They showed that the perception of synthetic vowels was affected by the formant structure of the vowels in a preceding synthetic carrier sentence. They suggested that all of the vowels spoken by a given talker contain "personal information" (anatomical and physiological features related to vocal tract shape) inherent in the talker's voice, and that this information, in part, determines the perceptual quality of each of the following vowels. Some researchers have referred to this issue in terms of listeners "compensating" for the source and/or vocal tract characteristics of the talker (Fourcin, 1968; Rand, 1971; Summerfield, 1975; Summerfield & Haggard, 1973). Other researchers have focused more narrowly on the perception of vowels and have suggested that listeners "recalibrate" or "rescale" the vowel space as a function of the voice of the talker (e.g. Bladon, Henton, & Pickering, 1984; Dechovitz, 1977; Disner, 1980; Gerstman, 1968; Nearey, 1978; Syrdal & Gopal, 1986; Verbrugge, Strange, Shankweiler, & Edman, 1976). Vowel normalization algorithms may have a basis in the neurophysiology of the human auditory system (Sussman, 1986).

In the last few years, a small handful of perceptual studies have reported that changes from trial to trial or from stimulus to stimulus in the voice of the talker affect the perception of both vowels and consonants. Using an identification task, Verbrugge, Strange, Shankweiler, and Edman (1976) showed that the identification of natural vowels was more accurate when the vowel stimuli were drawn from tokens produced by a single talker than when the stimuli were drawn from a variety of talkers including men, women, and children (see also Assman, Nearey, & Hogan, 1982; Weenink, 1986). Apparently, a change in the voice of the talker from trial to trial interfered in some manner with the perceptual processing and encoding of the vowels (see, however, Strange, Verbrugge, Shankweiler, & Edman, 1976 for conflicting results). Changes in perception have also been shown to occur with consonants when the talker varies from trial to trial (Fourcin, 1968).

In addition to changes in perceptual identification, processing time also appears to be affected by changes in the voice of the talker. In an early study on this problem, Summerfield and Haggard (1973) demonstrated that latencies for categorizing synthetic vowels were slower when target items were preceded by syllables designed to acoustically emulate a different voice (see also Summerfield, 1975). The authors suggested that the increase in response time due to talker variability reflected some additional processing time needed for vocal tract normalization to be carried out on the input speech signal. According to Summerfield and Haggard, the perceptual system appears to "retune" itself on the basis of vocal tract characteristics each time it encounters an item produced by a different talker.

Variability or uncertainty about a talker's voice has also been found to affect perceptual processing time in a same-different matching task (Allard & Henderson, 1975; Cole, Coltheart, & Allard, 1974). Cole et al. (1974) demonstrated that response latencies to auditory "same" judgments were slower when the voice of two target words differed. Thus, taken together, there appears to be some experimental evidence in the literature to suggest that, at least at the segmental acoustic-phonetic level, variability in the voice characteristics of the talker has reliable perceptual consequences for human listeners in a variety of perceptual tasks.

The results of these studies are consistent with the idea that changes in perceptual performance due to variability or uncertainty about the talker reflect the operation of some type of general perceptual normalization process operating at an early acoustic-phonetic level of analysis in speech perception (see also, Sussman, 1986). However, perceptual processing at this level constitutes only a small portion of the processing involved in the perception of fluent speech (McClelland & Elman, 1986; Pisoni & Luce, 1987). At this time, there is little research available in the literature on whether perceptual effects due to talker variability are also present at the lexical level. Current models of spoken word recognition (Forster, 1976, 1979; Klatt, 1979; Luce, 1986; Marslen-Wilson, 1987; Marslen-Wilson & Tyler, 1980; McClelland & Elman, 1986; Morton, 1969, 1982) have little, if anything, to say about the potential importance that talker voice information may have with regard to the recognition of spoken words. Since the possible effects of acoustic differences due to the talker on word recognition are not addressed in these models, one may be led to believe that the perceptual effects due to talker variability are confined to early, pre-lexical levels of processing and have little impact on the recognition of spoken words or subsequent comprehension processes which are typically assumed to occur at higher, more abstract levels of analysis.

Indeed, in current word recognition models much emphasis is placed on factors such as word frequency and lexical structure. Studies examining the effects of word frequency on word recognition (e.g. Grosjean, 1980; Howes & Solomon, 1951; Morton, 1969; Savin, 1963; Scarborough, Cortese, & Scarborough, 1977; Solomon & Postman, 1952; Stanners, Jastrzembski, & Westbrook, 1975) and, more recently, the effects of lexical structure (Eukel, 1980; Landauer & Streeter, 1973; Luce, 1986) have repeatedly demonstrated robust effects of these factors on word recognition performance using a variety of experimental paradigms. Based on these findings, researchers developing models of spoken word recognition have explicitly incorporated mechanisms into their models to account for the perceptual effects of frequency and lexical structure. It is interesting to note that while emphasis has been placed on these factors other potential variables such as talker variability that may also affect word recognition have received little if any attention. If a factor such as talker variability has equally consistent and substantial effects on word recognition as word frequency and lexical structure, it should also be treated with the same importance in models of spoken word recognition and incorporated in theoretical discussions of speech perception.

There is one study in the literature demonstrating that talker variability may have significant effects on spoken word recognition. Creelman (1957) conducted an intelligibility study in which he investigated the effects of talker variability on the recognition of spoken PB (phonetically-balanced) words. Creelman presented lists of monosyllabic words in noise to a group of five listeners. The words were presented in lists consisting of words spoken by one, two, four, eight, or sixteen talkers. The results showed that the words presented in the lists spoken by two or more talkers were identified less accurately than words presented in the list spoken by only a single talker. The differences in performance were relatively small, on the order of 7--10%. Creelman suggested that these results reflected relatively "minor" adjustments made by the perceptual system. Unfortunately, Creelman used a relatively small set of words and provided little in the way of any theoretical discussion of the results and their impact on spoken word recognition.

Creelman's study provides a starting point from which to further investigate the effects of talker variability on spoken word recogntion. However, in order to properly assess the importance of talker variability on word recognition, the effects must be assessed in conjunction with other variables known to produce substantial effects on performance under a variety of experimental conditions. As a result, we also examined the effects of word frequency and lexical density (a measure related to the structure and distribution of words in the lexicon). By studying the effects of talker variability along with these other variables, we hoped to obtain evidence demonstrating that talker variability is an important factor in speech perception that must be incorporated into current conceptions of spoken word recognition.

In order to determine whether talker variability produces substantial effects on spoken word recognition, experimental procedures must be used that are appropriate for investigating word recognition and lexical access. The perceptual studies examining talker variability effects that were reviewed above, with the exception of Creelman (1957), all involved perceptual tasks that emphasized the perception of acoustic cues in nonsense syllables. In order to generalize these earlier results, we used perceptual identification and naming tasks with familiar spoken words. These two tasks are suited to measuring perceptual performance at a point after which word recognition has already occurred, thus insuring that response decisions will be made on the basis of the identity of the word and not on the acoustic cues or segments contained in the stimulus.

In the first experiment, we attempted to replicate the findings of Creelman (1957) using a similar experimental procedure with a larger set of highly familiar words. In this experiment, talker variability and lexical density were manipulated. Talker variability was manipulated by having listeners identify, in one condition, words produced by a single talker or, in a second condition, words produced by fifteen different talkers. Stimulus items were selected to differ in lexical density, a measure related to the perceptual similarity of words in the mental lexicon. Landauer and Streeter (1973) and Eukel (1980) originally reported that lexical structure affects word recognition and lexical access and that high- and low-frequency words differ in a variety of ways above and beyond just frequency of occurrence in the language. More recently, using a variety of auditory and visual perceptual tasks, Luce (1985, 1986) has found that structural factors, including lexical density, were important determinants of word recognition performance. Lexical density was defined in the present experiment as the number of words differing from a given lexical item by one phoneme substitutions (see Greenberg & Jenkins, 1964). Using this simple distance metric, words could be indexed with regard to the composition of their similarity "neighborhoods". High-density words are words that have a large number of acoustically similar neighbors, whereas low-density words are words that have a much smaller number of phonetically confusable neighbors. Words of high lexical density and low lexical density were selected in order to study the perceptual effects of this variable and how it interacts with talker variability. Luce (1986) has shown that low-density items are identified more accurately and faster than high-density items because there are fewer confusable items in low-density similarity neighborhoods.

Several outcomes are possible. First, if talker variability has detrimental effects on spoken word recognition performance, then recognition accuracy should be worse under conditions where subjects received stimuli from many talkers compared to only one talker (i.e, mixed-talker versus single-talker conditions). Second, overall performance should differ as a function of lexical density. Low-density items should be identified correctly more often than high-density items. Finally, the use of a perceptual identification procedure involves the presentation of words in a background of white noise at different signal-to-noise ratios. Words should be identified correctly more often at high S/N ratios compared to low S/N ratios.

Experiment 1

Method

Subjects. Thirty-seven undergraduate students from introductory psychology courses at Indiana University volunteered to be subjects. Fifteen subjects served as talkers to produce the stimulus materials and another 22 subjects served as listeners in the perceptual experiment. Each subject participated in one 1-hour session and received partial course credit for the experiment as part of a requirement in introductory psychology. All subjects were native speakers of English and reported no history of a speech or hearing disorder at the time of testing.

Stimulus Materials. The stimuli consisted of 68 spoken words obtained from each of fifteen different talkers. All talkers had a midwestern dialect. The test items consisted of CVC monosyllabic English words containing a wide variety of consonants (i.e. stops, fricatives, affricates, liquids, and nasals) and vowels. Each talker's utterances were recorded on audiotape in a sound-attenuated booth (IAC Model 401A) using an Electro-Voice Model D054 microphone and a Crown 800 series tape recorder. Each stimulus item appeared on a CRT screen in front of the subject, embedded in the carrier sentence "Say the word ___ for me", where the blank corresponded to a particular target word. The talker was instructed to read the entire sentence aloud in a normal voice at a constant speaking rate. Utterances were recorded from seven male talkers and eight female talkers. The carrier sentences were subsequently low-pass filtered at 4.8 kHz and then converted to digital form via a 12-bit analog-to-digital converter using a 10 kHz sampling rate. The target words were then digitally edited from the carrier sentences to produce the final experimental materials used in the study. RMS amplitude levels among words were digitally equated using a software package designed to modify speech waveforms.

An on-line lexical database based on Webster's Pocket Dictionary (Webster's Seventh Collegiate Dictionary, 1967) was used to compute measures of lexical density. This database was used to compute a distance measure for each stimulus based on neighborhood similarity (see Luce, 1985, 1986). The measure of lexical density used in selecting these words was defined as the number of words (neighbors) differing by one phoneme from the stimulus that a particular word had in the lexicon. Low-density words were selected to have a value of ten or less; high-density words were selected to have a value of 15 or greater. Thirty-four words were selected for each condition, resulting in a total of 68 test stimuli. In addition, raw word frequency estimates were obtained for each word from the Kucera and Francis (1967) word count. The mean overall frequency counts for the low and high-density items was 41.8 and 54.2, respectively. A one-way ANOVA was conducted on the low and high-density items using word frequency as the variable. The results showed that the low-density words and the high-density words did not significantly differ from each other in word frequency ($F[1,66] = 0.35$, $p > .5$).

The final constraint used in selecting words was related to their subjective familiarity. Familiarity ratings on a scale from one (unknown) to seven (familiar and well-known) were obtained for the words in the database from subjects in a previous study (Nusbaum, Pisoni, & Davis, 1984). The stimuli selected for the present study met a 95% criterion of familiarity. All 68 stimuli were rated at 6.65 or above on the familiarity rating scale. Thus, all target words were rated as highly familiar by subjects. This manipulation insured that subjects were familiar with the words used in the experiment and that the items were, with very high probability, in the subjects' mental lexicon.

Procedure. Three experimental factors were manipulated: Talker variability, lexical density, and signal-to-noise (S/N) ratio. Talker variability was manipulated as a between-subjects factor. Subjects in the single-talker group listened to words from the same talker throughout the test session, while subjects in the mixed-talker group listened to words drawn from all fifteen talkers. Each group contained eleven subjects. In the single-talker group, each subject received the 68 stimuli produced by one of the fifteen different talkers. That is, each subject received stimuli from a different talker. This procedure minimized the possibility that inherent intelligibility differences between talkers would confound any effects due to talker variability displayed between the two groups. In the mixed-talker group, five words were randomly selected for presentation from each of the eight female talkers and four words were selected from each of the seven male talkers. The manipulation of lexical density created two within-subject conditions: high-density and low-density.

Finally, signal-to-noise ratio was manipulated to vary the level of performance. Each word was presented at three different S/N ratios: +10 dB, 0 dB, and -10 dB. Each subject received each word at each S/N ratio. For all three S/N conditions, the background noise remained constant at 70 dB SPL while the signal level was presented at 80 dB SPL, 70 dB SPL, and 60 dB SPL for the three conditions.

The experimental procedure employed an auditory perceptual identification task. Each stimulus item was embedded in noise and presented to subjects binaurally over matched and calibrated TDH-39 headphones. For each trial, subjects were instructed to identify the word that was presented and then type their response on a CRT terminal. A prompt appeared on the CRT screen immediately after presentation of the stimulus to indicate that a response should be initiated. Subjects were instructed to type in an English word corresponding to what they thought they had heard on each trial. Subjects were not given any information about what words to expect during the experiment except that they would all be familiar English words. After all subjects responded, a message appeared on the CRT indicating that the next stimulus would be presented. Subjects were not given any feedback concerning the correct response after each trial. A two-second ISI occurred between presentation of the message and the next trial.

Three separate blocks of 68 trials were run. A two-minute rest period occurred between each block. Each test word was presented once in each block and each test word was presented at a different S/N ratio in each particular block. Within a block, words occurred at all three of the S/N ratios so that one-third of the words were presented at each S/N ratio in each block. The assignment of S/N ratio to each word, as well as presentation of words within each block, was randomized. Stimulus output and data collection were controlled on-line by a PDP-11/34a computer. Stimuli were output via a 12-bit digital-to-analog converter at a 10 kHz sampling rate and were low-pass filtered at 4.8 kHz before presentation through the headphones.

Results and Discussion

The data were scored for percent correct identification of the target words. In Figure 1 and in Table 1, the identification results are displayed for the single and mixed-talker conditions for high and low lexical-density words at each of the three S/N ratios examined in the experiment.

Insert Figure 1 and Table 1 about here

A four-way ANOVA was carried out on the arcsine transformed data (see reference note 2). The factors in the design were talker variability (single or mixed-talker), density (high or low), S/N ratio (+10, 0, or -10) and block (1st, 2nd, or 3rd block of trials). Three significant main effects were obtained. First, there was a significant effect of talker variability ($F[1,20] = 7.9$, $p < .02$). Identification was more accurate for the single-talker condition compared to the mixed-talker condition (40.6% correct and 33.9%, respectively, averaged over all conditions). This result demonstrates that a change in the talker's voice from trial to trial does, in fact, produce detrimental effects on spoken word recognition in this perceptual identification task.

Second, as expected, a main effect was found for S/N ratio ($F[2,40] = 838.0$, $p < 0.01$). Identification performance was most accurate in the +10 S/N condition, less accurate in the 0 S/N condition, and least accurate in the -10 S/N condition (63.6%, 42.2%, and 5.9% correct, respectively). Thus, performance varied reliably as a function of the discriminability of the speech signal.

Third, a main effect of test block was observed ($F[2,40] = 30.8$, $p < .01$). Performance in the first block of trials was less accurate than the second and third block (32.9%, 40.0%, and 39.9% correct, respectively). Newman-Keuls posthoc tests showed that performance in the second and third block did not differ reliably while performance in the first block was significantly different from the other two. This result suggests, not surprisingly, that experience with the stimuli and experimental procedures obtained in the first block led to better performance in the later blocks.

Finally, no significant main effect of lexical density was obtained ($F[1,20] = 1.9$, $p > .2$). Although the results were in the expected direction (36.5% and 38.0% correct, respectively, for high-density and low-density words), the differences obtained in this study were not large enough to reach statistical significance. In addition, no significant interactions were obtained.

The results of this experiment provide important new data concerning the effects of talker variability on spoken word recognition. The finding that performance was substantially worse in the mixed-talker condition compared to the single-talker condition demonstrates that changes from trial to trial due to talker variability have detrimental effects on the processes involved in recognizing spoken words. In the mixed-talker condition, it appears likely that some type of perceptual readjustment or normalization related to processing a talker's voice was made on each trial in order to facilitate the recognition of each test item. Clearly, the uncertainty of the trial-to-trial

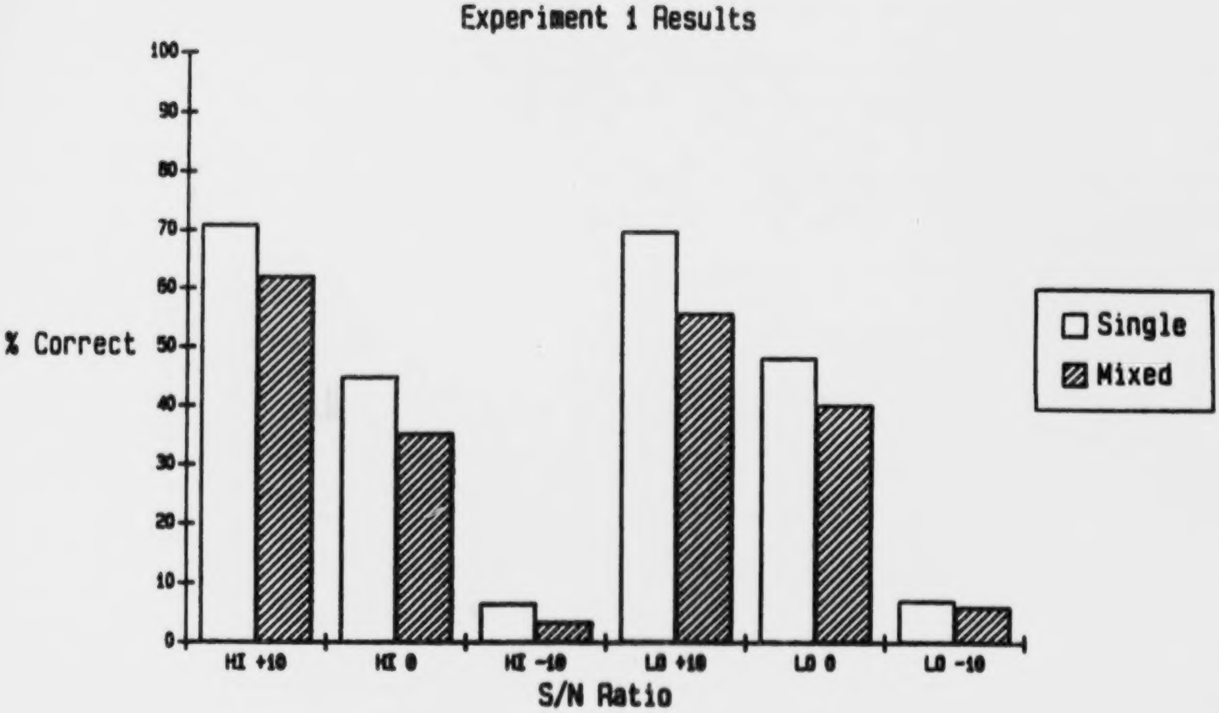


Figure 1. Overall mean percent correct performance collapsed over subjects for Experiment 1. Performance is shown for single and mixed-talker conditions as a function of high- and low-density words and S/N ratio.

Table 1

Mean overall percent correct identification performance in Experiment 1 for single and mixed-talker groups as a function of lexical density and S/N ratio.

		Density	
	S/N ratio	High	Low
Single Talker	+10	66.5	70.0
	0	45.0	48.3
	-10	6.6	7.2
Mixed Talker	+10	62.1	55.9
	0	35.3	40.2
	-10	3.5	6.2

variability in the talker's voice affects identification performance for the same set of items. The only difference between the conditions was the context in which the test items were presented. These results replicate the results of the earlier study conducted by Creelman (1957) using different words under similar conditions.

With regard to the effects of lexical density, we failed to find a significant effect as we had originally predicted. Although density does have reliable and systematic effects on spoken word recognition as reported by Luce (1986), under the conditions of the present experiment, these effects were not large and did not reach a statistically significant level. However, our results were in the expected direction (36.5% and 38.0% correct, respectively, for high-density and low-density words).

In summary, the results of the first experiment demonstrate that talker variability produces substantial effects on the perception of spoken words degraded by noise. These results also suggest that talker variability may be an important factor that has been ignored in current models of word recognition. Unfortunately, the use of the perceptual identification task does not permit an assessment of the effects of talker variability and lexical variables on perceptual processing time. In addition, the use of the perceptual identification task does not reveal whether talker variability and lexical density affect the perception of stimuli that are not degraded by noise. Because of these considerations, a second experiment was conducted to examine the effects of talker variability using a naming task. A number of researchers have used the naming procedure to examine effects of variables related to word recognition and lexical access because it provides a method of collecting latency data along with identification responses to stimuli uncorrupted by noise (Balota & Chumbley, 1984, 1985; Luce, 1986). Thus, in using this procedure, the effects of talker variability and lexical density can be assessed for words presented in the clear.

Experiment 2

Method

Subjects. Twelve undergraduate students from an introductory psychology course at Indiana University served as subjects. Each subject participated in two 1-hour sessions that were conducted on two consecutive days. Each subject received partial course credit for participating in the experiment. All subjects were native speakers of English and reported no history of a speech or hearing disorder at the time of testing.

Stimulus Materials. The same stimuli used in Experiment 1 were used for the present experiment. All aspects of the stimuli remained exactly the same. The stimuli in the low-density and high-density conditions were equated for manner class of the initial consonant so that an equal number of stimuli containing initial stops, strong fricatives, weak fricatives, nasals, liquids, and semivowels could be assigned to each condition to reduce measurement variability.

Procedure. Two within-subject experimental factors were manipulated, talker variability (single versus mixed) and lexical density (high versus low). The talker and lexical density conditions were the same as in Experiment 1. Items in the single-talker condition were drawn from one talker and items in the mixed-talker condition were drawn from fifteen different talkers. Each subject received the single-talker condition on one day of testing and the mixed-talker condition on the other day of testing. The conditions were counterbalanced across subjects. Each subject was run individually in a small testing booth containing headphones, a microphone, and a CRT monitor.

The experimental procedure consisted of requiring subjects to name words aloud as fast and as accurately as they could. Each stimulus was binaurally presented over TDH-39 headphones to the subject at a listening level of 75 dB. The subject was required to initiate a vocal naming response after hearing each target word. Subjects were instructed to repeat the target word into a voice-activated microphone (Electro-Voice Model D054) as soon as they could identify the word. They were instructed to keep their lips approximately four inches from the microphone. The distance was monitored by an experimenter during the course of the experiment. A message appeared on the CRT in front of each subject after each response was collected, indicating that the next stimulus would be presented. Each stimulus item was presented two seconds after collection of the response.

Four blocks of 68 trials were run on each day. A two-minute rest period occurred between each block. Each stimulus item was presented once within each block. The order of stimulus presentation within a block was randomized. Half of the subjects received the single-talker condition on the first day of testing and half received the mixed-talker condition on the first day. An experimenter sat near the subject during the experiment and monitored a CRT screen that displayed the target words for each trial. The experimenter listened to the subject's vocal response and compared it to the correct target word for the trial displayed on the experimenter's monitor. After each vocal response, the experimenter hit one of two keys on the computer to indicate whether the vocal response for that particular trial was "correct" or "incorrect". An incorrect response was defined as any vocal response in which the word was mispronounced or consisted of a word other than the correct target word. If a word was mispronounced, the item was returned to the pool of items for a block so it could be presented again. Stimulus presentation and data collection were controlled on-line by a PDP-11/34A computer as in Experiment 1.

Results and Discussion

The data were analyzed in terms of overall percent correct identification and response latencies. Response latencies were analyzed for correct responses only. The response latencies are considered first. Table 2 shows the mean latencies collapsed over subjects for the single and mixed-talker conditions for high and low-density words.

Insert Table 2 about here

A three-way ANOVA was conducted on the mean latency data. The factors in the design were talker variability, lexical density, and trial block. A significant effect of talker variability was found ($F[1,11] = 10.7$, $p < .01$). Response latencies were faster for the words in the single-talker condition than for the same words in the mixed-talker condition (608.4 and 678.3 msec, respectively). A significant effect of trial block was also observed ($F[3,33] = 5.3$, $p < .01$). Response latencies decreased as a function of practice over

Table 2

Mean response latency (msec) in Experiment 2 for correct responses for single and mixed-talker conditions as a function of lexical density.

	Density	
	High	Low
Single-talker	611.2	605.7
Mixed-talker	677.2	679.4

blocks (673.9, 640.7, 635.0, and 623.7 msec, respectively, for blocks one through four). However, Newman-Keuls post-hoc tests revealed that performance did not differ significantly between each of the blocks of trials. No main effect of lexical density and no significant interactions were obtained.

Overall percent correct collapsed over subjects is displayed in Table 3 for the single and mixed-talker conditions for high and low-density words.

Insert Table 3 about here

A three-way ANOVA was also conducted on the arcsine transformed identification data. The factors were talker variability, lexical density, and trial block. A significant effect of talker variability was obtained ($F[1,11] = 7.4$, $p < .02$). Identification performance was better for words in the single-talker condition compared to words in the mixed-talker condition, (95.8% and 91.4% correct, respectively), replicating the results of the first experiment. No other significant main effects were found in this analysis. The only significant interaction was density x block ($F[3,33] = 6.3$, $p < .01$). Newman-Keuls tests revealed that high-density items were identified correctly more often than low-density items in the third block of trials. This interaction was due to a crossover in identification accuracy between high and low-density conditions over blocks.

Overall, the effects due to talker variability found in the first experiment were replicated in this study using a naming paradigm in which the stimulus items were not degraded by noise. Performance as measured by identification and latencies was consistently worse in the mixed-talker condition compared to single-talker condition. These results provide additional evidence that talker variability from trial to trial not only affects overt identification responses but also affects the time course of perceptual processing. Taken together, the results from the first two experiments demonstrate that changes from trial to trial in the talker's voice, at least within the perceptual identification and naming paradigms, result in reliable effects on spoken word recognition. The context that the test items are presented in appears to reliably affect identification and response time.

With regard to the manipulation of lexical density, as in Experiment 1, we found no significant main effect of density on response latencies or on identification responses. However, density entered into an interaction with trial block for identification responses only. Although an examination of the interaction revealed that high-density words were identified correctly more often than low-density words in one block of trials, this effect was only significant in one out of four blocks of trials. Since there were no main effects of density on identification responses or response latencies, we will ignore the one significant interaction with block.

One can think of a number of possible reasons why the lexical density manipulation may not have produced any reliable effects in the present experiment and in the previous one. First, the procedures used to compute lexical density may have been too crude. The use of one-phoneme substitutions may not be the best procedure to compute similarity neighborhoods. It is possible that a metric based on specific phoneme confusions may be more

Table 3

Mean overall percent correct identification in Experiment 2 for single and mixed-talker conditions as a function of lexical density.

	Density	
	High	Low
Single-talker	96.6	95.0
Mixed-talker	91.8	91.1

appropriate (see Luce, 1986). Using this alternative method, Luce (1986) observed significant effects of lexical density on spoken word recognition.

Second, Luce (1986) has shown that along with lexical density, factors such as acoustic-phonetic confusibility (derived from phonetic confusion matrices), word frequency, and mean neighborhood frequency also have independent effects on spoken word recognition. Although word frequency was controlled in each density condition, acoustic-phonetic confusibility and neighborhood frequency were not. Thus, it is possible that variations in these factors may have obscured any systematic effects of lexical density on performance in these two experiments.

Finally, we used a relatively small number of test stimuli in this experiment and they were all very highly familiar monosyllablic CVC words. It is possible that the processing of items with these characteristics may differ from items exhibiting a wider range of acoustic-phonetic diversity and subjective familiarity (see Luce, 1986).

Since the effects of lexical density may be difficult to reveal, at least under the present conditions with these stimuli, the manipulation of a different variable related to lexical processing may help us to understand the effects of talker variability on spoken word recognition and may provide some insight into the relative impact that these factors have on spoken word recognition. One such variable that has been extensively investigated in the word recognition literature is word frequency (Morton, 1969). The effects of word frequency on various perceptual processes have been documented using a wide variety of experimental paradigms (e.g. Grosjean, 1980; Howes & Solomon, 1951; Morton, 1969; Savin, 1963; Scarborough et al., 1977; Solomon & Postman, 1952; Stanners et al., 1975). This variable has been shown to produce large and reliable effects in most word recognition tasks. Generalized across a large number of studies, high-frequency words are typically perceived faster and more accurately than low-frequency words. Given that word frequency effects are extremely robust, an investigation of this variable may provide additional information about the effects of talker variability on spoken word recognition.

In the next experiment, talker variability and word frequency were manipulated in a naming paradigm similar to that used in Experiment 2. The effects of talker variability should be shown by a pattern of performance similar to that obtained in Experiment 2. In addition, if word frequency has a signficant effect on performance, responses to high-frequency words should be faster than responses to low-frequency words.

Experiment 3

Method

Subjects. Seventy undergraduate students from an introductory psychology course at Indiana University served as subjects. Fifty subjects participated in one 1-hour session that was devoted to screening stimuli for the experiment. Twenty additional subjects participated in one 1-hour session for the experiment proper. Each subject received partial course credit for their participation. All subjects were native speakers of English and reported no history of a speech or hearing disorder at the time of testing.

Stimulus Materials. The stimuli consisted of 96 naturally spoken words obtained from each of seven male and eight female talkers of a midwestern dialect. These stimuli were English monosyllabic and polysyllabic words drawn from the Modified Rhyme Test (House et al., 1965) and the Phonetically Balanced word lists (Egan, 1948). The recording and editing of the stimuli were conducted in a manner similar to that previously described for Experiments 1 and 2. The test words differed in word frequency as defined by the Kucera and Francis (1967) frequency counts. Low-frequency items were defined as those words with values of 10 or fewer occurrences per million in the Kucera and Francis count; high-frequency items were defined as those words with values of 100 or above per million. Forty-eight low-frequency words and 48 high-frequency words were selected for use in this experiment.

As in the two previous studies, the stimuli were all rated as highly familiar by subjects (above 6.65 on the seven-point scale) using the norms obtained in an earlier study. A one-way ANOVA was run on the low and high-frequency items to assess differences in lexical density. The results showed that the high- and low-frequency words did not differ significantly in density. The mean number of neighbors was 22.1 and 21.4 for high-frequency and low-frequency items, respectively.

The stimuli used in the present experiment were further screened to insure that the items distributed across the single and mixed-talker conditions did not differ in intelligibility. A total of 1440 stimuli (96 from each of 15 talkers) were presented to fifty subjects for identification in a separate experiment. The experimental procedure was a word identification task. Words were presented in the clear and subjects were required to type in a string of characters corresponding to the word they heard. Five groups of ten subjects were run. Each group was presented with stimuli from three different talkers. All stimulus items displayed scores of 90% or above correct identification on this test.

Items from one male talker were selected for use in the single-talker condition. Items drawn from all 15 talkers were selected for use in the mixed-talker condition. Seven words were drawn from six talkers and six words were drawn from nine of the talkers. Intelligibility scores were equated between the talker conditions so that each stimulus in the mixed-talker condition possessed the same score as the corresponding identical stimulus in the single-talker condition.

Procedure. Two experimental factors, talker variability and word frequency, were manipulated in a completely within-subjects design. Each subject received both high- and low-frequency items from both single-talker and mixed-talker conditions in the testing session. The experimental procedure consisted of the naming task, which was conducted in the same fashion as described for Experiment 2. Each subject received two blocks of 96 trials in which high- and low-frequency words were randomly presented once within each block. One block consisted of words from the single-talker condition while the other block consisted of words drawn from the mixed-talker condition. The order of blocks was counterbalanced across subjects.

Results and Discussion

The data were analyzed separately for percent correct identification and response latencies for correct responses only. Table 4 shows mean latencies collapsed over subjects for the single and mixed-talker conditions as a function of word frequency.

Insert Table 4 about here

A two-way ANOVA was conducted on the latency data to assess the effects of talker variability and word frequency. A significant main effect of talker variability was obtained ($F[1,19] = 11.1$, $p < .01$). Response latencies were faster in the single-talker condition compared to the mixed-talker condition (834.2 and 868.9 msec, respectively). A significant main effect of frequency on response latency was not obtained ($F[1,19] = 2.2$, $p > .15$), although response latencies were slightly faster for high-frequency words than low-frequency words (847.1 and 856.0 msec, respectively). The interaction of frequency and talker variability also was not significant ($F[1,19] = 1.3$, $p > .26$).

For identification performance, the mean percent correct identification scores averaged over subjects is shown in Table 5 as a function of talker variability and frequency.

Insert Table 5 about here

A two-way ANOVA was conducted on the arcsine transformed identification data to assess the effects of talker variability and word frequency. A significant main effect was observed for talker variability ($F[1,19] = 38.3$, $p < .01$). Identification performance was better in the single-talker condition compared to the mixed-talker condition (97.8% and 92.9% correct, respectively). In addition, a significant main effect of word frequency was also found ($F[1,19] = 21.5$, $p < .01$). High-frequency words were identified more accurately than low-frequency words (97.2% and 93.5% correct, respectively). The interaction of talker variability and frequency was not significant.

The results of the present experiment replicate and extend the results obtained in our earlier experiments. First, a robust effect of talker variability was observed for both dependent variables. Faster response latencies and more accurate identification performance was found in the single-talker condition compared to the mixed-talker condition using a within subjects design. Thus, the effects of talker variability were replicated using a larger number of stimuli which were explicitly controlled for intelligibility in isolation.

Second, as reported in other studies, word frequency had an effect on overall identification performance. However, word frequency did not affect response times in the naming task. High-frequency words were correctly identified more often than low-frequency words and response latencies for high-frequency words were slightly faster than low-frequency words, although this difference was not significant. Frequency-related differences in identification performance, although significant, were relatively small (see Table 5); both high and low-frequency words were identified at fairly high levels of accuracy. One explanation for the absence of frequency effects on

Table 4

Mean overall response latency (msec) in Experiment 3 for correct responses for single and mixed-talker conditions as a function of word frequency.

	Word Frequency	
	High	Low
Single-talker	825.6	842.9
Mixed-talker	868.7	869.2

Table 5

Mean overall percent correct identification performance in Experiment 3 for single and mixed-talker conditions as a function of word frequency.

	Word Frequency	
	High	Low
Single-talker	99.2	96.5
Mixed-talker	95.3	90.6

response latency may be the use of the naming task. Balota and Chumbley (1984) found that frequency effects, although significant, were substantially reduced when a naming task was used. They argued that the effects of word frequency may be more salient when using experimental procedures that tap later stages of processing where subject biases may operate.

The results of this experiment indicate that the effects of talker variability on spoken word recognition are at least as substantial, if not more so, than the effects of word frequency. The effects on identification accuracy of these variables were of approximately the same magnitude (see Table 5). However, a large and significant effect of talker variability on response latencies was also obtained, indicating that talker variability affected processing time as well. Given the possibility mentioned earlier that frequency effects may be reduced in the naming task, we decided to conduct a fourth experiment that employed a perceptual identification task that was similar to the one used in Experiment 1. In using this task, the effects of word frequency may be more salient than those observed using the naming task.

In addition to examining the effects of talker variability and word frequency on perceptual identification, we also examined a factor related to the ease of encoding of the input signal. This factor involved degradation of the acoustic information using a novel signal processing technique (see Horrii, House, & Hughes, 1971; Salasoo & Pisoni, 1985). If the digital signal is degraded by randomly deleting samples of the original speech waveform, then the early auditory processes involved in extracting information relevant to phonetic distinctions should be affected because the initial acoustic cues are degraded. By manipulating the degree of degradation, the relative effects of talker variability and word frequency can be examined. This method of degradation was chosen over alternative methods, such as imposing a uniform background of white noise over the stimulus, because any effects due to degradation are a direct consequence of physical disruption and/or distortion of the original information in the signal. That is, the stimulus information that is presented is not degraded by masking noise.

Talker variability was manipulated in a fashion similar to that of the previous experiments. Word frequency was also manipulated in order to assess whether the well-known effects of word frequency could be replicated using a perceptual identification procedure with the same stimuli.

Experiment 4

Method

Subjects. Thirty undergraduate Indiana University students with the same qualifications described earlier were used as subjects. Each subject participated in one 1-hour session and received partial course credit for their participation.

Stimulus Materials. The 96 stimuli used in Experiment 3 provided the basis for the stimuli used in the present study. These stimuli were modified by degrading the speech signal using digital signal processing techniques. The technique used to produce the degraded signals involved a computer program which flips the sign of the amplitude value of the digital waveform for each sample at randomly determined points over a specified proportion of the waveform. For example, stimuli at the 10% degradation level consisted of the original stimulus with 10% of the amplitude values at random points having values opposite to those contained in the original digital file. Degrading the stimuli in this manner resulted in utterances in which a percentage of the acoustic information deleted was simply replaced by noise. This resulted in stimuli which were intelligible but sounded somewhat "noisy" or "distorted".

Three sets of 96 stimuli were used. Each set consisted of items degraded at one level. The degradation levels were specified at 10%, 20%, and 30% of the waveform. Except for these changes, all other aspects of the stimuli remained the same as in Experiment 3.

Procedure. The three experimental factors manipulated were talker variability, word frequency, and percent degradation level. Word frequency and degradation level were manipulated within subjects while talker variability was manipulated between subjects by using two separate groups. In the single-talker group, subjects received stimuli from one male talker as in Experiment 3; in the mixed-talker group, subjects received stimuli drawn from all 15 male and female talkers. The experimental procedure consisted of a perceptual identification task that was the same as the one used in Experiment 1. Each subject listened to a stimulus word and typed in a string of characters as a response on a computer terminal corresponding to the word that he/she thought was presented. Subjects were told that the stimuli that they would be presented with would sound "noisy" or "distorted" and they were to pay close attention to the words and try to identify them as best as they could even if they had to guess.

Subjects were presented with three blocks of 96 stimuli in which the high- and low-frequency items were randomly presented within each block. Degradation level was blocked, such that each block of trials contained stimuli at one degradation level only. The order of blocks was counterbalanced across subjects by a latin square design. Stimuli were presented at a comfortable listening level of 75 dB.

Results and Discussion

The responses were analyzed in terms of percent correct identification. Table 6 shows the mean overall percent correct identification averaged over subjects for the single and mixed-talker groups as a function of word frequency and signal degradation level.

Insert Table 6 about here

A three-way ANOVA was conducted on the arcsine transformed identification data for the main variables, talker variablility, word frequency, and degradation level. As expected, a significant main effect of talker variability was observed ($F[1,28] = 91.6$, $p < .01$). Identification performance was better for the single-talker condition compared to the mixed-talker condition (69.1% and 48.1% correct, respectively). A main effect of word frequency was also found ($F[1,28] = 161.9$, $p < .01$). High-frequency words were identified more accurately than low-frequency words (64.3% and 52.8% correct, respectively). Finally, a significant main effect of degradation was also obtained ($F[2,56) = 91.7$, $p < .01$). Performance became worse as the degradation level increased (75.4%, 56.9%, and 43.5% correct, respectively, for 10%, 20%, and 30% degradation). Newman-Keuls post-hoc tests

Table 6

Mean overall percent correct identification performance in Experiment 4 for single and mixed-talker conditions as a function of signal degradation level and word frequency.

		Word Frequency	
	Degradation Level	High	Low
Single Talker	10%	87.8	74.6
	20%	76.1	63.2
	30%	66.8	46.0
Mixed Talker	10%	73.1	66.1
	20%	46.5	41.8
	30%	35.7	25.4

revealed that performance differed reliably between all three degradation levels used in the experiment.

Insert Figure 2 about here

A number of significant two-way interactions were also obtained. First, the interaction of talker variability and degradation was significant (F[2.56] = 7.3, p < .01). Figure 2 shows performance as a function of talker and degradation level. As can be seen in the figure, performance decreased more for items from the mixed-talker condition compared to the single-talker condition when the degradation level increased from 10% to 20%. However, the residual difference across talker groups remained about the same between the 20% and 30% levels. Newman-Keuls tests revealed that performance between single and mixed-talker conditions was significantly different at each degradation level and that performance within each talker condition was significantly different between degradation levels.

Insert Figure 3 about here

A significant interaction between degradation and frequency was also obtained (F[2,56] = 10.5, p < .01). Figure 3 shows performance for the frequency conditions as a function of degradation level. As shown here, the differences in performance between high- and low-frequency items remained about the same at the 10% and 20% degradation levels, but the differences became greater at the 30% level. Newman-Keuls tests showed that performance between high- and low-frequency words was significantly different at each degradation level, and that performance between degradation levels differed significantly within each frequency condition.

Insert Figure 4 about here

Finally, a significant interaction between talker variability and word frequency was obtained (F[1,28] = 14.4, p < .01). Figure 4 shows performance for the talker conditions as a function of word frequency. As shown here, the difference in performance between high- and low-frequency words was greater for the single-talker condition than for the mixed-talker condition. Newman-Keuls tests showed that performance between single and mixed-talker conditions differed significantly for both high- and low-frequency words. Newman-Keuls tests also showed that within each talker condition performance was significantly different between high- and low-frequency words.

The results of Experiment 4 provide further evidence of reliable and robust effects of talker variability on spoken word recognition. As in our previous experiments, perceptual performance was worse when the stimulus items

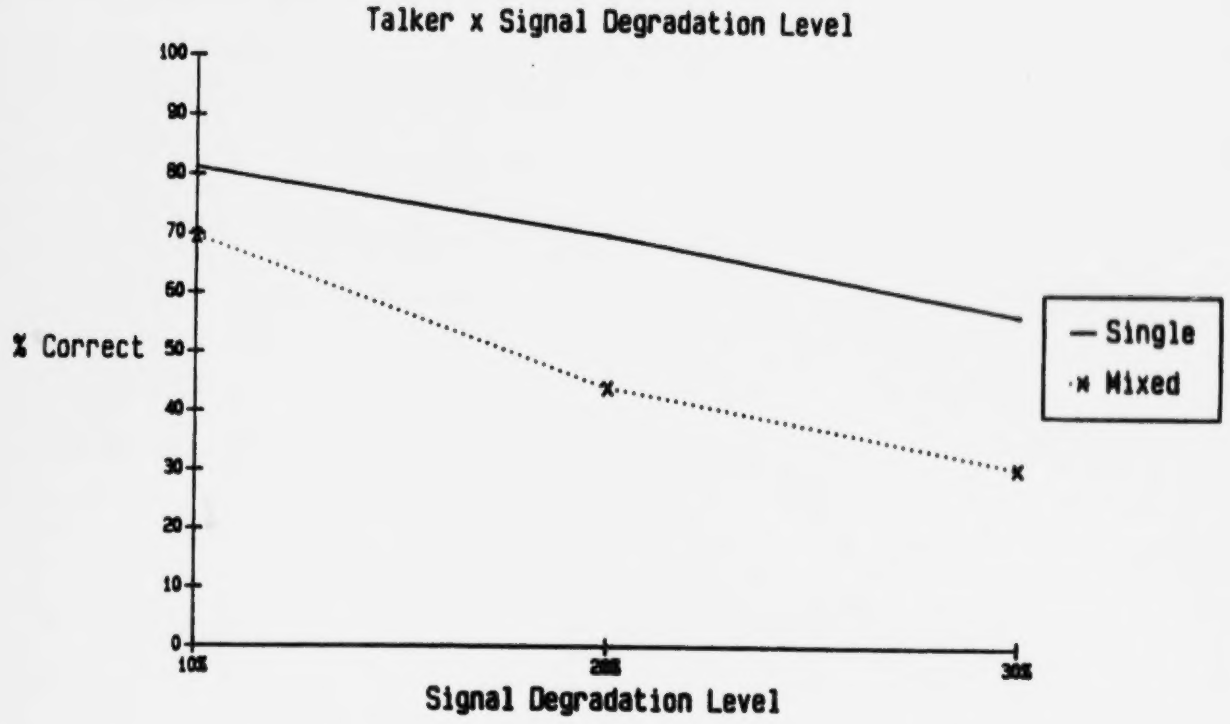


Figure 2. Mean percent correct performance collapsed over subjects and word frequency conditions for Experiment 4. Performance is shown for single and mixed-talker groups as a function of signal degradation level.

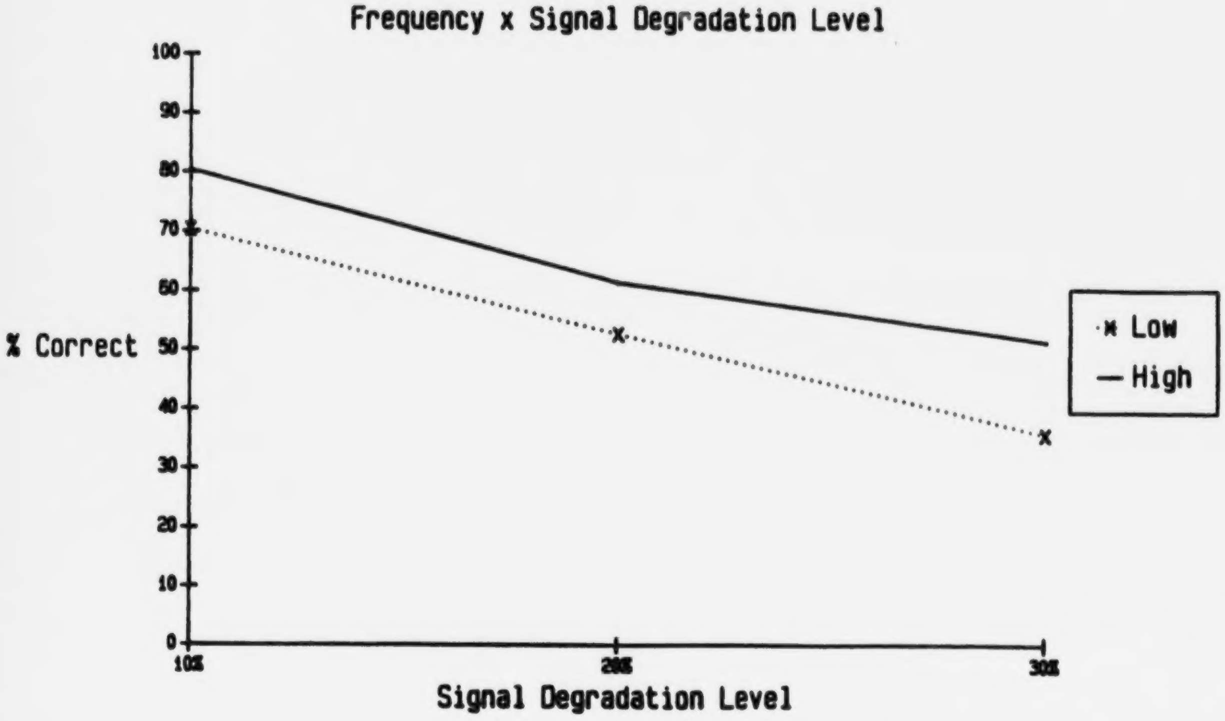


Figure 3. Mean percent correct performance collapsed over subjects and talker conditions for Experiment 4. Performance is shown for high- and low-frequency words as a function of signal degradation level.

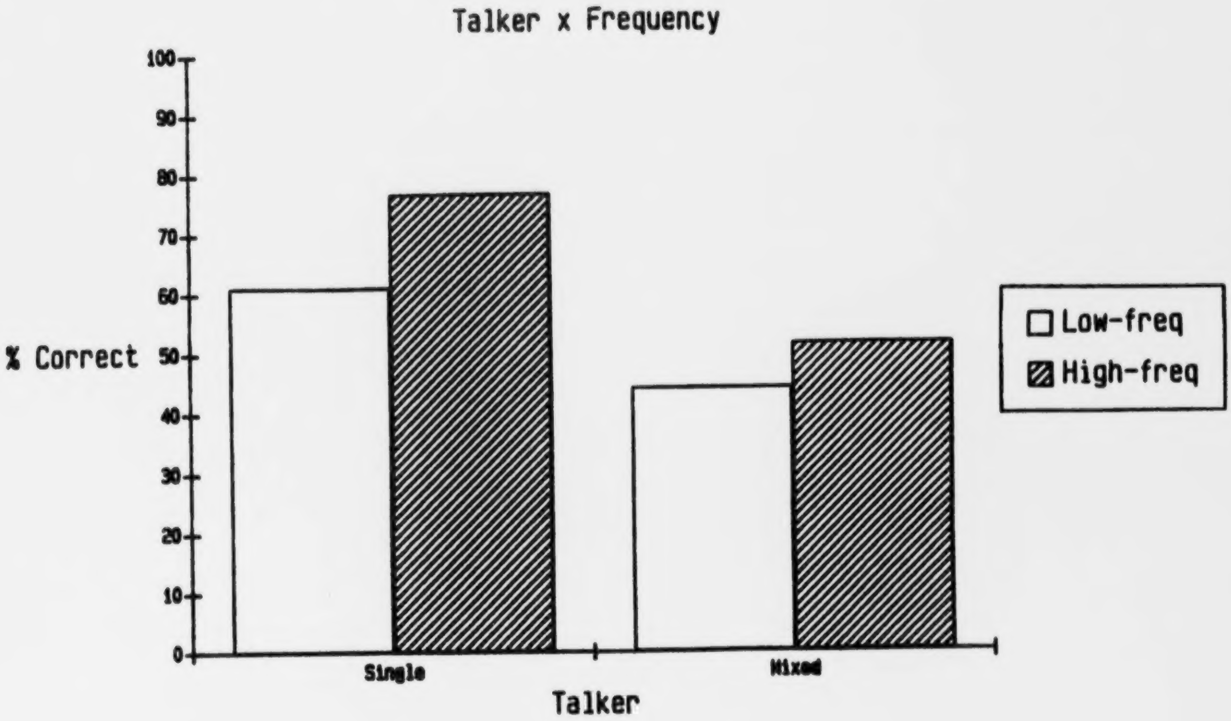


Figure 4. Mean percent correct performance collapsed over subjects and signal degradation levels for Experiment 4. Performance is shown for single and mixed-talker conditions as a function of word frequency.

were produced by different talkers on each trial than when they were produced by a single talker. This effect appears to be extremely consistent from experiment to experiment across different sets of stimulus materials and different tasks.

A significant effect of word frequency was also observed in the present experiment. The effects of word frequency appear to be more salient using a perceptual identification paradigm compared to a naming paradigm, a finding that was reported by Balota and Chumbley (1985). In addition, inspection of the data for the word frequency manipulation reveals that the magnitude of the effects are approximately the same as the magnitude of the effects produced by talker variability. Thus, it appears that talker variability and word frequency both have substantial effects on spoken word recognition, at least in the context of the present study.

A number of interactions between the variables were also obtained. An examination of the interaction of talker variability and signal degradation shows that as the degree of degradation of the signal increased from 10% to 20%, performance became worse for items in the mixed-talker condition compared to the same items in the single-talker condition. This result indicates that when the processing of low-level acoustic cues in the signal becomes increasingly disrupted as a result of signal degradation, the processing of talker-specific information also becomes impaired. The finding is consistent with a view suggesting that talker normalization is intimately related to processes involved in encoding the sensory input in the speech signal into a phonetic representation.

An interaction also occurred between degradation level and word frequency. The decrease in performance for high-frequency words compared to low-frequency words was about the same at the 10% and 20% degradation levels, but the difference became larger at the 30% level (see Figure 3). One account of this interaction is that the result may simply be due to "guessing" or response bias. When the words become extremely degraded, such as at the 30% degradation level, subjects may be more likely to guess high-frequency words than low-frequency words in making a response (see Goldiamond & Hawkins, 1958; Luce, 1986). This hypothesis is supported by the observation that there is little difference in performance between high and low-frequency words at the 10% and 20% levels, but a much larger difference at the 30% level. Only 43.5% of the test words were correctly identified overall at this level of degradation.

The final interaction to be considered was between talker variability and word frequency. The pattern of results indicated that the differences in performance between high- and low-frequency words were greater for the single-talker condition than the mixed-talker condition. An explanation for this pattern of results is unclear, although it may be related to the methods used to create the distortion in the stimulus materials or the amount of active rehearsal given to an item at the time of initial encoding (see Martin, Mullennix, Pisoni, & Summers, 1987). There is no immediately obvious reason why word frequency, particularly when it is viewed as a form of response bias, should have larger effects on items produced by a single talker which are presumably encoded more efficiently than stimuli produced by mixed talkers. Typically, word frequency manipulations produce greater effects when the information specifying the items is ambiguous or degraded. We have no explanation of this curious result at the present time although it may be reflecting some underlying difference caused by talker variability.

General Discussion

Taken together, the results of the present set of experiments have implications for models of spoken word recognition and previous accounts of perceptual normalization in speech perception. First, the effects of talker variability on spoken word recognition performance observed in the present study suggest that the processes involved in speech perception apparently include some mechanism or set of mechanisms that adjust for differences in a talker's voice and these mechanisms have a processing "cost" associated with them. When the voice of the talker is changed from trial to trial, perceptual processing of highly familiar CVC words becomes impaired. Isolated words are identified less accurately and require more processing time for recognition. Based on results obtained in both perceptual identification and naming tasks, we suggest that some resource-demanding mechanism is used by the listeners to compensate for the physical differences in the stimuli produced by different talkers. It is important to emphasize here that the speech waveforms were always identical across the two conditions we examined. The only differences were in the context in which the items were presented to the listeners.

Second, the results of the present study indicate that talker variability is an important factor that must be considered in models of word recognition and lexical access and integrated into current theoretical descriptions. Our results demonstrated repeatedly, under a variety of experimental conditions, that talker variability produces substantial and reliable effects on the processes involved in recognizing spoken words. When comparing these effects to the effects of word frequency and lexical density, two measures that have been shown to have substantial effects on word recognition, the effects of talker variability appear to be more robust and less dependent on the particular task. We obtained significant effects of talker variability on both identification and processing time, while we did not obtain any significant effects of lexical density and we obtained significant effects of word frequency only on overt identification responses. Word frequency, and to a lesser extent, lexical structure, have been typically given a great deal of importance in the development of models of word recognition and lexical access (e.g., Forster, 1976, 1979; Luce, 1986; Morton, 1969). More recently, researchers have begun to pay more attention to acoustic-phonetic factors and their involvement in spoken word recognition (Luce, 1986; Marslen-Wilson, 1987; Marslen-Wilson & Tyler, 1980; Pisoni & Luce, 1987). At the very least, our results suggest that the relationship of talker normalization to the processes involved in word recognition and lexical access should be further investigated and the findings integrated in models of word recognition and lexical access.

There are two possible ways in which talker variability in the present set of experiments may have produced its effects. First, as we mentioned earlier in the introduction, the results of a number of studies concerned with vowel and consonant perception demonstrate that changes in a talker's voice affect processes at an early segmental acoustic-phonetic level (Assman et al., 1982; Fourcin, 1968; Rand, 1971; Strange et al., 1976; Summerfield, 1975; Summerfield & Haggard, 1973; Verbrugge et al., 1976; Weenink, 1986). With regard to spoken word recognition, it is possible that the effects of talker variability we obtained in the present study are due to perceptual processes and operations that are confined to an analysis of early segmental information in the speech waveform. The output of these processes consists of a more abstract canonical representation that is passed on to higher-level processes related to word recognition, with the perceptual deficits arising at an early acoustic-phonetic level "cascading" up the system. Thus, talker normalization processes may be related to other low-level sensory encoding processes which are also sensitive to changes and variability in acoustic information in the speech signal.

Indeed, one factor that produces large and reliable acoustic changes in the signal is variations in speech rate. In a series of studies, J.L. Miller and her colleagues have provided extensive evidence that the phonetically relevant acoustic properties of the signal are not extracted in an absolute manner, but instead, are processed with regard to the rate at which the speech was produced (see Miller, 1981; 1986; 1987). Miller has proposed that the processing of speech rate information occurs at a relatively early stage of speech processing (Miller, Green, & Schermer, 1984). In one recent study, Miller et al. (1984) demonstrated that when the effects of semantic context on word identification are eliminated, substantial effects due to speech rate information still remain. This result is consistent with the proposal offered by Miller (1987) that speech rate normalization occurs at a fairly early level of processing independent of processes related to the analysis of semantic information. On the basis of other experimental work, Miller (1987) argues that the processing of rate information is "mandatory" and takes place within a "phonetic module" which analyzes and interprets the information in the speech signal in terms of phonetic qualities. Given the evidence which supports the hypothesis that rate normalization processes occur at a relatively early level, it seems plausible, and perhaps even quite likely, that talker normalization processes may also operate at an early stage of perceptual analysis. This hypothetical stage makes use of processes involved in the acoustic-phonetic analysis of the speech signal into abstract phonetic categories and representations needed to access words in long-term memory (see Pisoni & Luce, 1987).

One result reported in Experiment 4, that the effects of talker variability become greater when the acoustic information in the speech signal becomes more physically degraded, is consistent with the idea that talker normalization processes are intimately related to early encoding processes that produce a phonetic representation of the signal. Because the perceptual processes affected by methods of signal degradation are precisely those which extract auditory and/or phonetic featural information from the acoustic signal, it is possible that talker normalization processes occur predominantly at an early acoustic-phonetic level and not later on at more abstract stages associated with word recognition or lexical access.

Another way in which talker variability may have produced effects on performance in the present study involves the idea that talker-specific perceptual features are actually retained for short periods of time in higher-level representations of the input that are matched to words in the lexicon. By this account, the effects of talker variability do not arise entirely from earlier acoustic-phonetic levels of processing, but instead are due to interactions caused by the presence of talker-specific properties in the lexical matching process. This account would incorporate the notion that talker-specific features from a previous input item or items remain in memory and produce interference when a subsequent item is perceived. Although there is some evidence in the literature that talker-related features are retained in long-term memory (Craik & Kirsner, 1974; Geiselman & Belezza, 1976, 1977) and may cause interference (Martin et al., 1987; Mullennix & Pisoni, 1987), it is not clear how these features may be represented, nor is it clear by what manner they would produce interference with subsequent items. In some cases, it is even possible for talker variability to produce improved performance in serial recall tasks (see Logan & Pisoni, 1987).

The results obtained in the present series of experiments are consistent with results of previous research concerning the effects of talker variability on perception at the segmental acoustic-phonetic level using isolated vowels and CV nonsense syllables (Assman et al., 1982; Fourcin, 1968; Rand; 1971; Strange et al., 1976; Summerfield, 1975; Summerfield & Haggard, 1973; Verbrugge et al., 1976; Weenink, 1986). The most consistent finding from the present series of experiments was that word recognition was affected strongly and consistently by changes in a talker's voice from item to item. This result suggests that the processes operating on voice information apparently incur a processing debt even in simple tasks such as perceptual identification and naming. Whether talker normalization is a relatively simple and automatic "vocal tract normalization" process (Summerfield & Haggard, 1973) or involves more complex perceptual adjustments remains to be investigated (Liberman & Mattingly, 1985). Although much recent research suggests that the perception of vowels may be accomplished by algorithmic rescaling or recalibration processes (Bladon et al., 1984; Dechovitz, 1977; Gerstman, 1968; Nearey, 1978; Syrdal & Gopal, 1986; but, see Disner, 1980), the perception of connected, fluent speech produced by different talkers obviously involves much more than simple rescaling of differences in static vocal tract configurations. Differences in dynamic articulatory trajectories resulting from non-linear control of the articulators as well as differences in glottal waveforms and numerous other factors known to differ between talkers all lead to different acoustic realizations of an utterance. Thus, talker normalization processes in speech perception may be much more complex and dynamic than previously described in earlier work using simple isolated vowels and CV nonsense syllables. For the present, however, our results using isolated, highly familiar words are consistent with the findings reported earlier in the literature using isolated vowels and nonsense syllables.

Obviously, further research will be necessary to understand the precise nature of the differences between talkers and to further characterize the nature of the perceptual mechanisms responsible for talker normalization effects. This work will need to examine further the relationship between talker normalization processes and the other perceptual processes involved in developing a segmental representation of the input signal for spoken word recognition. In addition, more research will need to be conducted in order to more clearly determine the relationship of talker variability to the processes involved in spoken word recognition and lexical access. Other research from our laboratory using memory and selective attention paradigms (e.g. Martin et al., 1987; Nusbaum, Greenspan, & Pisoni, 1986) has already provided additional evidence concerning the effects of talker variability on speech perception and spoken word recognition. In the past, most studies of speech perception have used only a single talker throughout an entire experiment. The present results using multiple talkers demonstrate robust and reliable differences due to talker variability in processing the same signals. These findings will need to be incorporated into current theoretical conceptions of speech perception and spoken language processing.

References

Allard, F., & Henderson, L. (1976). Physical and name codes in auditory memory: The pursuit of an analogy. *Quarterly Journal of Experimental Psychology*, *28*, 475-482.

Assman, P.F., Nearey, T.M., & Hogan, J.T. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America*, *71*, 975-989.

Balota, D.A., & Chumbley, J.I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology*, *10*, 340-357.

Balota, D.A., & Chumbley, J.I. (1985). The locus of word-frequency effects in the pronunciation task: lexical access and/or production? *Journal of Memory and Language*, *24*, 89-106.

Bladon, R.A., Henton, C.G., & Pickering, J.B. (1984). Towards an auditory theory of speaker normalization. *Language and Communication*, *4*, 59-69.

Carr, P.B., & Trill, D. (1964). Long-term larynx-excitation spectra. *Journal of the Acoustical Society of America*, *36*, 2033-2040.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*, (pp. 254-259). Hillsdale, N.J.: Erlbaum.

Cole, R.A., Coltheart, M., & Allard, F. (1974). Memory of a speaker's voice: Reaction time to same- or different-voiced letters. *Quarterly Journal of Experimental Psychology*, *26*, 1-7.

Craik, F.I.M., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*, *26*, 274-284.

Creelman, C.D. (1957). Case of the unknown talker. *Journal of the Acoustical Society of America*, *29*, 655.

Crowder, R.G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception and Psychophysics*, *5*, 365-373.

Dechovitz, D. (1977). Information conveyed by vowels: a confirmation. *Haskins Laboratories Status Report on Speech Research*, SR-51/52, 213-219.

Disner, S.F. (1980). Evaluation of vowel normalization procedures. *Journal of the Acoustical Society of America*, *67*, 253-261.

Egan, J.P. (1948). Articulation testing methods. *Laryngoscope*, *58*, 955-991.

Eukel, B. (1980). A phonotactic basis for word frequency effects: Implications for automatic speech recognition. *Journal of the Acoustical Society of America*, *68*, S33.

Fant, G. (1973). Speech sounds and features. Cambridge, MA: MIT Press.

Forster, K.I. (1976). Accessing the mental lexicon. In R.J. Wales and E. Walker (Eds.), New approaches to language mechanisms. Amsterdam: North-Holland.

Forster, K.I. (1979). Levels of processing and the structure of the language processor. In W.E. Cooper and E.C.T. Walker (Eds.), Sentence processing: Psycholinguistic studies presented to Merrill Garrett. Hillsdale, N.J.: Erlbaum.

Fourcin, A.J. (1968). Speech-source interference. IEEE Transactions on Audio and Electroacoustics, ACC-16, 65-67.

Geiselman, R.E., & Bellezza, F.S. (1976). Long-term memory for speaker's voice and source location. Memory and Cognition, 4, 483-489.

Geiselman, R.E., & Bellezza, F.S. (1977). Incidental retention of speaker's voice. Memory and Cognition, 5, 658-665.

Gerstman, L. (1968). Classification of self-normalized vowels. IEEE Transactions on Audio and Electroacoustics, ACC-16, 78-80.

Goldiamond, I., & Hawkins, W.F. (1958). Vexierversuch: The logarithmic relationship between word-frequency and recognition obtained in the absence of stimulus words. Journal of Experimental Psychology, 56, 457-463.

Greenberg, J.H., & Jenkins, J.J. (1964). Studies in the psychological correlates of the sound system of American English. Word, 20, 157-177.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. Perception and Psychophysics, 28, 267-283.

House, A.S., Williams, C.E., Hecker, M.H.L., & Kryter, K.D. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. Journal of the Acoustical Society of America, 37, 158-166.

Horri, Y., House, A.S., & Hughes, G.W. (1971). A masking noise with speech envelope characteristics for studying intelligibility. Journal of the Acoustical Society of America, 49, 1849-1856.

Howes, D.H., & Solomon, R.L. (1951). Visual duration threshold as a function of word probability. Journal of Experimental Psychology, 41, 401-410.

Joos, M.A. (1948). Acoustic phonetics. Language, Suppl. 24, 1-136.

Klatt, D.H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. Journal of Phonetics, 7, 279-312.

Kucera, F., & Francis, W. (1967). Computational analysis of present day American English. Providence, RI: Brown University Press.

Ladefoged, P. (1980). What are linguistic sounds made of?. Language, 56, 485-502.

Ladefoged, P., & Broadbent, D.E. (1957). Information conveyed by vowels. Journal of the Acoustical Society of America, 29, 98-104.

Landauer, T.K., & Streeter, L.A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. Journal of Verbal Learning and Behavior, 12, 119-131.

Logan, J. & Pisoni, D.B. (1987). Talker variability and the recall of spoken word lists: A replication and extension. Research on Speech Perception Progress Report No. 13. Bloomington, IN: Indiana University.

Luce, P.A. (1985). Similarity neighborhoods and word frequency effects in visual word identification: Sources of facilitation and inhibition. Research on Speech Perception Progress Report No. 11. Bloomington, IN: Indiana University.

Luce, P.A. (1986). Neighborhoods of words in the mental lexicon. Research on Speech Perception Technical Report No. 6. Bloomington, IN: Indiana University.

Marslen-Wilson, W.D. (1987). Functional parallelism in spoken word-recognition. In U.H. Frauenfelder and L.K. Tyler (Eds.), Spoken word recognition. Cambridge, MA: MIT Press.

Marslen-Wilson, W.D., & Tyler, L.K. (1980). The temporal structure of spoken language understanding. Cognition, 8, 1-71.

Martin, C.S., Mullennix, J.W., Pisoni, D.B., & Summers, W.V. (1987). Effects of talker voice information on recall of spoken word lists. Research on Speech Perception Progress Report No. 13. Bloomington, IN: Indiana University.

Mattingly, I.G., Studdert-Kennedy, M., & Magen, H. (1983). Phonological short-term memory preserves phonetic detail. Journal of the Acoustical Society of America, 73, B6.

McClelland, J.L., & Elman, J.L. (1986). Interactive processes in speech perception: the TRACE model. In J.L McClelland and D.E. Rumelhart (Eds.), Parallel distributed processing, vol. 2: Psychological and biological models. Cambridge, MA: MIT Press.

Miller, J.L. (1981). Effects of speaking rate on segmental distinctions. In P.D. Eimas and J.L. Miller (Eds.), Perspectives on the study of speech. Hillsdale, NJ: Erlbaum.

Miller, J.L. (1986). Rate-dependent processing on speech perception. In A. Ellis (Ed.), Progress in the psychology of language, vol. III. Hillsdale, NJ: Erlbaum.

Miller, J.L. (1987). Mandatory processing in speech perception. In J.L. Garfield (Ed.), *Modularity in knowledge representation and natural-language understanding*. Cambridge, MA: Erlbaum.

Miller, J.L., Green, K., & Schermer, T.M. (1984). A distinction between the effects of sentential speaking rate and semantic congruity on word identification. *Perception and Psychophysics*, *36*, 329-337.

Monsen, R.B., & Engebretson, A.M. (1977). Study of variations in the male and female glottal wave. *Journal of the Acoustical Society of America*, *62*, 981-993.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*, 165-178.

Morton, J. (1982). Disintegrating the lexicon: An information processing approach. In J. Mehler, E. Walker, and M. Garrett (Eds.), *On mental representation*. Hillsdale, N.J.: Erlbaum.

Mullennix, J.W., & Pisoni, D.B. (1987). Talker variability and processing dependencies between word and voice. *Research on Speech Perception Progress Report No. 13*. Bloomington, IN: Indiana University.

Nearey, T.M. (1978). *Phonetic feature systems for vowels*. Published by Indiana University Linguistics Club, Bloomington, IN.

Nusbaum, H.C., Greenspan, S.L., & Pisoni, D.B. (1986). Perceptual attention in monitoring natural and synthetic speech. *Research on Speech Perception Progress Report No. 12*. Bloomington, IN: Indiana University.

Nusbaum, H.C., Pisoni, D.B., & Davis, C.K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception Progress Report No. 10*. Bloomington, IN: Indiana University.

Peterson, G.E., & Barney, H.L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*, 175-184.

Pisoni, D.B., & Luce, P.A. (1987). Acoustic-phonetic representation in word recognition. In U.H. Frauenfelder and L.K. Tyler (Eds.), *Spoken word recognition*. Cambridge, MA: MIT Press.

Rand, T.C. (1971). Vocal tract size normalization in the perception of stop consonants. *Haskins Laboratories Status Reports on Speech Research*, SR-25/26, 141-146.

Salasoo, A., & Pisoni, D.B. (1985). Interaction of knowledge sources in spoken word identification. *Journal of Memory and Language*, *24*, 210-231.

Savin, H.B. (1963). Word-frequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America*, *35*, 200-206.

Scarborough, D., Cortese, C., & Scarborough, H. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 1-17.

Solomon, R.L., & Postman, L. (1952). Frequency of usage as a determinant of recognition thresholds for words. Journal of Experimental Psychology, 43, 195-201.

Stanners, R.F., Jastrzembski, J.E., & Westbrook, A. (1975). Frequency and visual quality in a word-nonword classification task. Journal of Verbal Learning and Verbal Behavior, 90, 45-50.

Strange, W., Verbrugge, R.R., Shankweiler, D.P., & Edman, T.R. (1976).

Consonant environment specifies vowel identity. Journal of the Acoustical Society of America, 60, 213-224.

Summerfield, Q. (1975). Acoustic and phonetic components of the influence of voice changes and identification times for CVC syllables. Report of Speech Research in Progress, 2(4). The Queen's University of Belfast, Belfast, Ireland.

Summerfield, Q., & Haggard, M.P. (1973). Vocal tract normalisation as demonstrated by reaction times. Report on Research in Progress in Speech Perception, No. 2. The Queen's University of Belfast, Belfast, Ireland.

Sussman, H.M. (1986). A neuronal model of vowel normalization and representation. Brain and Language, 28, 12-23.

Syrdal, A. K., & Gopal, H.S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. Journal of the Acoustical Society of America, 79, 1086-1100.

Verbrugge, R.R., Strange, W., Shankweiler, D.P., & Edman, T.R. (1976). What information enables a listener to map a talker's vowel space?. Journal of the Acoustical Society of America, 60, 198-212.

Webster's Seventh Collegiate Dictionary (1967). (Library Reproduction Service, Los Angeles).

Weenink, D.J.M. (1986). The identification of vowel stimuli from men, women, and children. Proceedings 10 from the Institute of Phonetic Sciences of the University of Amsterdam, 41-54.

Reference Notes

1. We use the expression "changes in a talker's voice" throughout the manuscript to refer to the variability in the production of specific test items spoken by *different* talkers. While the term is potentially ambiguous, we are concerned primarily in this research with variability between talkers rather than variability within a specific talker.

2. All data analyses reported for the identification data in all experiments were performed on nonlinear arcsine transformations of the raw identification data (see Cohen & Cohen, 1975). The arcsine transformation was defined as

$$A = 2 \text{ arcsine} \sqrt{p},$$

where p is a proportion and A is a transformed value (measured in radians).

[RESEARCH ON SPEECH PERCEPTION Progress Report No. 13 (1987) Indiana University]

Effects of Talker Variability on Recall of Spoken Word Lists*

C. S. Martin, J. W. Mullennix, D. B. Pisoni, and W. V. Summers

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, IN. 47405

---

*This research was supported, in part, by NIH Research Grant NS-12179-11 and, in part, by NIH Training Grant NS-07134-09 to Indiana University.

Abstract

Previous perceptual studies have shown that trial-to-trial changes in the voice of a talker have perceptual consequences at both segmental and lexical levels of processing. In order to investigate the effects of talker variability on recall, four list-learning experiments were conducted using lists of monosyllabic English words spoken by either a single talker or different talkers. Serial recall of early list items was better for lists spoken by a single-talker than for lists spoken by multiple talkers. This result was not obtained in a free recall experiment. A third experiment utilized a memory preload procedure using visually-presented digits. Recall of the preload digits was superior when items in a subsequent list were spoken by a single talker compared to multiple talkers. A fourth experiment used a retroactive interference task to eliminate contributions of short-term memory on recall. The interference task did not reduce the differences in recall performance between talker conditions. The results of the first three experiments suggest that the encoding and rehearsal of spoken lists produced by multiple talkers requires greater processing resources than lists produced by a single talker. The results of Experiment 4 suggest that the superior serial recall of early list items for single-talker word lists is not due to retrieval processes that are independent of initial perceptual encoding and subsequent rehearsal in short-term memory. Taken together, these experiments demonstrate that talker variability not only affects encoding processes at the time of input but also affects the efficiency of rehearsal processes used in transfering items into long-term memory.

Effects of Talker Variability on Recall of Spoken Word Lists

The acoustic properties of speech vary dramatically as a function of context, speaking rate, and a number of talker-related factors such as vocal tract configuration, glottal characteristics, vocal amplitude, and dialect. Many theorists have argued that in order for spoken language to be perceived rapidly and efficiently, some sort of perceptual process must compensate for the acoustic differences between individual talkers (e.g., Joos, 1948; Verbrugge, Strange, Shankweiler, & Edman, 1976). This perceptual compensation in speech perception suggests a form of perceptual constancy. Talker differences are thought to be "normalized" at fairly early stages of perceptual analysis so that linguistic units can be efficiently extracted from the speech waveform (Summerfield & Haggard, 1973). Although perceptual normalization has been recognized as an important research problem almost from the beginning of modern speech research, little is known about the nature of this type of perceptual compensation. Indeed, an examination of the published literature reveals that almost all research in speech perception that has used natural speech has employed stimulus tokens produced by a single talker. Human listeners rapidly perceive and understand speech signals produced by a wide variety of talkers and appear to display little, if any, additional effort or processing demands. The mechanisms used to perform these operations have not received much attention in the field of speech research.

However, some research relevant to this issue has been conducted. Several studies has shown that changes from stimulus to stimulus in talker voice affect vowel perception. Verbrugge et al. (1976) reported that vowel identification was superior for vowel stimuli produced by the same talker compared to vowel stimuli produced by different talkers. Summerfield and Haggard (1973) reported that synthetic vowels were categorized more slowly when they were preceded by synthetic syllables designed to acoustically emulate the voice characteristics of different talkers (see also Summerfield, 1975). Summerfield and Haggard (1973) suggested that the increase in response latencies reflected the processing time needed for a "vocal tract normalization" process, although they did not specify the nature of this process in any detail. Taken together, the results of these perceptual studies demonstrate that changes in talker voice have detrimental effects on processing at the segmental acoustic-phonetic level.

The effects of variability from item to item in the voice of a talker have also been examined at the lexical level. Creelman (1957) found that changes from word to word in the voice of a talker reduced identification performance for PB (phonetically balanced) words. Talker voice characteristics have also been found to affect response latencies in a same-different matching paradigm (Allard & Henderson, 1975; Cole, Coltheart, & Allard, 1974). Cole et al. (1974) reported that response latencies were slower for "same" judgements when the target words were produced by different talkers than when the items were produced by the same talker. More recently, several experiments in our laboratory have examined the effects of talker variability on spoken word recognition (Mullennix, Pisoni, & Martin, 1987). In a series of experiments, Mullennix et al. obtained results demonstrating detrimental effects on word recognition when talker voice changed from trial to trial compared to when the talker's voice remained the same across trials. Reliable effects of talker variability were obtained for both perceptual identification accuracy and response latency. Thus, it appears that changes in the voice of the talker produce perceptual deficits at the level of word recognition, as well as at earlier segmental levels.

Although earlier research has demonstrated that talker variability affects perceptual processing at segmental and lexical levels, little research has examined the effects of talker variability on the cognitive processes involved in memory. One study conducted by Craik and Kirsner (1974) examined the effects of talker variability on recognition memory. Subjects listened to spoken word lists in which the stimuli were produced by a male talker and a female talker. In a recognition memory test, list items were repeated in either the same voice or in a different voice from the one in which list items were originally presented. The results demonstrated that recognition of list items was faster and more accurate when words were repeated in the same voice as the original item. This facilitation due to talker voice remained constant over a 2-minute interval. Furthermore, subjects were able to accurately recall the voice in which words had originally been presented after a 2-minute lag. These results suggest that information about a talker's voice can be retained in memory for at least two minutes and that talker-specific features may be used to facilitate recognition memory for words.

One experiment has examined the effects of talker variability on the recall of words from memory. Mattingly, Studdert-Kennedy, and Magen (1983) examined the effects of changes in a talker's voice and dialect variation on serial recall for spoken word lists. Stimuli were spoken by either a single talker, three different talkers with the same dialect, or three different talkers with different dialects. The results indicated that recall performance for early list items was significantly worse when list items were produced by different talkers with different dialects compared to list items produced by a single talker or by three talkers with the same dialect. Mattingly et al. suggested that changes in dialect, but not in the voice of the talker within a dialect, affected encoding and/or rehearsal processes in memory, and that these effects were reflected in recall performance for early list items.

Several factors may have affected the outcome of the Mattingly et al. (1983) experiment. First, the use of only three talkers in the multiple-talker conditions may not have produced enough variability to demonstrate any reliable perceptual consequences of talker variability in this paradigm. It is possible that a wider range of variability in the voice of the talker may be required to exhibit such effects. Secondly, the stimulus items consisted of digit-names, which are a highly constrained and overlearned vocabulary. The use of digits as stimuli may have encouraged subjects to use rehearsal and retrieval strategies that are quite different from those used for a less constrained set of stimulus items. With highly constrained stimulus sets, subjects often engage in guessing strategies or other response strategies to improve their performance on a task (Miller, Heise, & Lichten, 1951).

Considering the possible problems with the Mattingly et al. (1983) experiment, and the paucity of research examining the effects of talker variability on memory processes, we felt that further research was needed to investigate the effects of talker variability on the recall of words. The present series of experiments investigated the effects of talker variability on recall of lists of isolated spoken words. This work follows from our earlier perceptual research showing reliable effects of talker variability on word identification tasks.

Recall performance can be used as an index of the capacity demands required for the encoding and rehearsal of different types of speech input (Luce, Feustel, & Pisoni, 1983). When given a list of isolated words to recall, subjects tend to recall more items from the first few and the last few positions in a list than from the middle positions of a list. The enhanced recall for early and late list items are known as the primacy effect and the recency effect, respectively, and have been well documented in the memory literature (see Crowder, 1976). In dual-process accounts of memory, primacy and recency effects are thought to reflect different memory stores (Atkinson & Shiffrin, 1968; Waugh & Norman, 1965), differences in depth of item processing (Craik, 1973), or differences in search accessibility (Shiffrin, 1970). Recency effects have been explained as reflecting the output of items from a short-term memory buffer (Glanzer, 1972; Waugh & Norman, 1965; but see Greene, 1986 for alternative explanations). Primacy effects, on the other hand, are explained as reflecting a greater number of rehearsals or more elaborative rehearsal devoted to early list items than to later list items. A number of theorists have suggested that a greater amount of rehearsal leads to a higher probability that an item will be transfered to long-term memory (Atkinson & Shiffrin, 1968; Bruce & Papay, 1970; Waugh & Norman, 1965). Alternatively, a greater amount of rehearsal may lead to stored images of greater strength, which are then more easily retrieved from memory (Shiffrin, 1970). There is a good deal of evidence that the amount of rehearsal devoted to early list items affects primacy recall performance (Baddeley & Hitch, 1977; Brodie & Prytulak, 1975; Rundis & Atkinson, 1970). Thus, primacy recall performance can be used as an index of the amount or type of rehearsal devoted to early list items.

It is now well-accepted that short-term memory is limited in its capacity to hold and process information (e.g., Shiffrin, 1976). Different amounts of processing resources will be available for a particular task, depending upon how much processing capacity is being allocated to other tasks. In a recall task, a limited amount of processing capacity is available for the encoding and rehearsal of stimulus items (Baddeley & Hitch, 1974). If the encoding of spoken word items produced by different talkers requires a greater amount of the limited-capacity resources in short-term memory, rehearsal processes should be less efficient for items in multiple-talker lists than for items in single-talker lists. Differences in the amount or efficiency of rehearsal for multiple-talker and single-talker word lists may therefore produce differences in primacy recall between these two conditions. Specifically, the presentation of multiple-talker lists may result in lower recall performance for early list items compared to single-talker lists. Thus, an examination of recall performance may provide a method to measure differences in the capacity demands required for the encoding and rehearsal of words spoken by either a single talker or by multiple talkers.

Several years ago, Luce, Feustel, and Pisoni (1983) used recall performance as an index of capacity demands for lists of naturally produced speech and synthetic speech. Luce et al. (1983) found that recall for synthetic word lists was worse than recall for naturally produced word lists at all serial positions within a list. In addition to this main effect of speech type, an interaction of natural/synthetic speech and serial position was also obtained. Differences in recall between the natural and synthetic lists were largest in the primacy region of the serial position curve. Luce et al. (1983) interpreted these results as support for the proposal that greater capacity demands are required for the encoding and subsequent rehearsal of synthetic speech compared to natural speech. Given these findings, one might expect that recall performance would be worse for lists spoken by multiple talkers compared to lists spoken by a single talker over all serial positions. These results would be expected if the processing of items in multiple-talker lists requires more processing resources in short-term memory than the processing of items in single-talker lists. Synthetic speech, however, is often misperceived by naive listeners; the differences in recall performance obtained by Luce et al. (1983) may reflect both a larger number of encoding errors as well as increased capacity demands for synthetic speech.

Lists of words produced by different talkers, on the other hand, should be perceived rapidly and efficiently by listeners because these words contain the redundant acoustic cues characteristic of natural speech. Performance decrements due to the encoding and rehearsal of list items may not be salient enough to affect recall for items in terminal list positions. The literature offers little evidence to motivate predictions about recall performance for single-talker and multiple-talker lists. Given the perceptual experiments demonstrating reliable effects of talker variability at the segmental and word levels, we were interested in determining whether talker variability would also have effects on recall of word lists and what the nature of these effects might be. It is possible that variability due to the voice of the talker in the multiple-talker lists will affect early encoding processes, with these effects cascading up the processing system to affect the rehearsal and transfer of items into long-term memory. On the other hand, talker variability may only affect encoding at early stages of perceptual analysis. These perceptual differences may then be encapsulated and not affect subsequent memory processes.

In the first experiment, serial recall of word lists containing 10 items was investigated. Word lists were constructed from items spoken by either a single talker, 10 talkers of the same gender, or five male and five female talkers. The two multiple-talker conditions were constructed to examine whether the increased talker variability due to gender differences would result in a greater effect on recall performance. Based on earlier work on the recall of synthetic speech (Luce, Feustel, and Pisoni, 1983), we predicted that recall performance for early list items would decrease as the amount of talker variability within a list increased. If recall differences in the primacy region of the serial position curve are obtained as a function of talker variability, this result would be consistent with the hypothesis that the processes involved in perceptual encoding and rehearsal require a greater amount of processing resources when there are changes in the voice of the talker from item to item within a list.

Experiment 1

Method

Subjects. Subjects were 112 undergraduate students at Indiana University who participated to fulfill a course requirement in introductory psychology. Each subject participated in one hour-long session. All subjects were native speakers of English who reported no history of a speech or hearing disorder at the time of testing.

Stimuli. The stimuli consisted of five lists of 10 monosyllabic English words. Words were originally recorded in isolation on audio tape and digitized via a 12-bit analog-to-digital converter on a PDP-11/34 computer. All word lists were generated from digital files stored in the computer. Three versions of each word list were prepared. In the single-talker lists, all list items were spoken by one talker. In the multiple-talker same-gender condition, the 10 list items were spoken by 10 different talkers of the same gender. In the multiple-talker different-gender condition, the 10 list items were spoken by five different male and five different female talkers.

Overall RMS amplitude levels for all words were digitally equated using a specialized signal processing package. Stimuli were low-pass filtered at 4.8 kHz and played to listeners through a 12-bit digital-to-analog converter over matched and calibrated TDH-39 headphones at 80 dB SPL. The presentation of the word lists was controlled by a PDP-11/34a minicomputer.

Words within a list were semantically unrelated, and differed from each other by at least two phonemes. All of the words used in the experiment had been previously tested for intelligibility in a seperate experiment using a different group of listeners. These items received identification scores of 95% correct or above when presented in isolation.

Procedure. Subjects were tested in groups of six or less in a sound-treated room. On each trial, subjects were presented with a spoken list of 10 words. They were then given 60 seconds to recall the words in the exact position in which they were presented. Subjects recorded their responses by printing them on a response sheet.

The inter-word interval for stimulus presentation was 1.5 seconds. Immediately before the presentation of each list, subjects heard a 500 ms 1000-Hz warning tone. Following presentation of each list, another tone signaled the end of the list and the beginning of the 60-second recall period. During this period, subjects were instructed to write down as many of the words as they could recall in the exact serial position in which they were presented. Subjects were told that items not recalled in the correct position would be scored as incorrect.

The talker variable was manipulated in a between subjects design. Subjects were randomly assigned to one of three conditions: single-talker, multiple-talker same-gender, or multiple-talker different-gender. Identical word lists were used in each condition; the conditions differed only in the voices used to produce the words. Each subject heard four blocks of the five lists of words for a total of 20 list presentations. The order of lists within each block and the order of stimuli within each list were randomized. Two practice lists were presented at the beginning of the experimental session in order to familiarize subjects with the experimental procedure.

Results and Discussion

Figure 1 shows the percentage of words correctly recalled as a function of serial position and talker condition averaged over all trials.

Insert Figure 1 about here

Inspection of Figure 1 indicates that the typical serial position curve was obtained for each of the three talker conditions. For each condition, recall performance for initial and final list items is better than recall of items in the middle of the list. In order to test for differences between the three talker conditions, a two-way ANOVA was conducted for the factors of

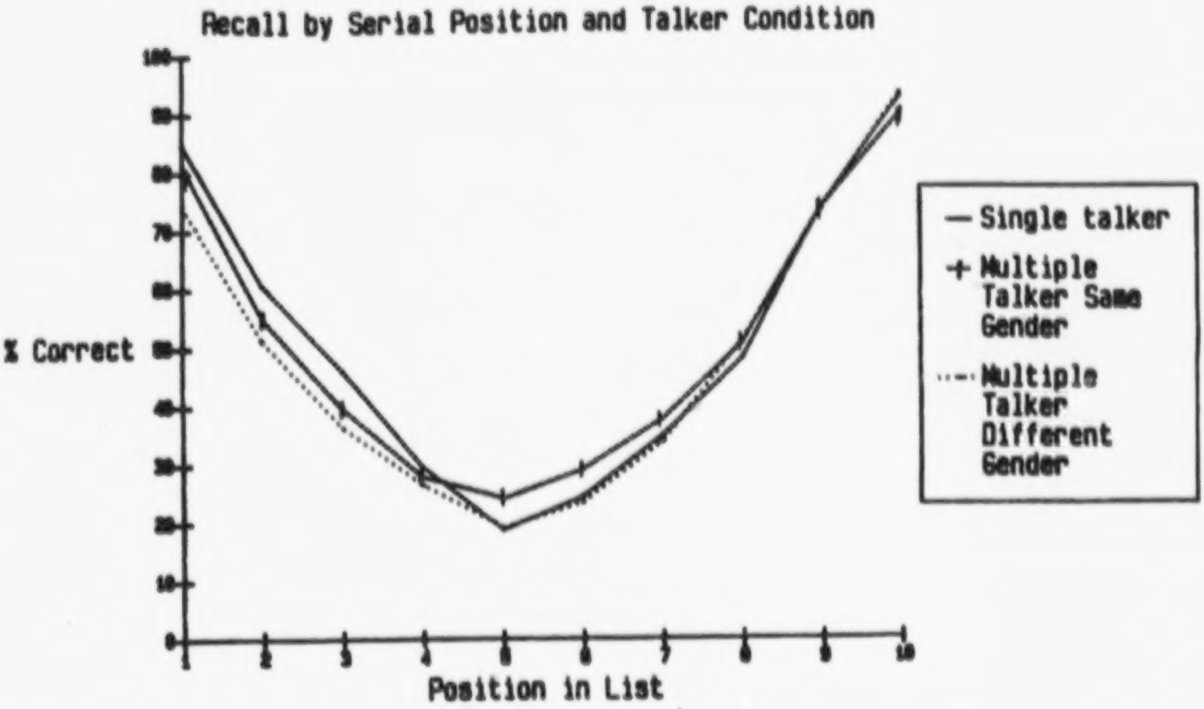


Figure 1. Mean percent correct serial recall collapsed over subjects as a function of serial position and talker condition for Experiment 1.

talker condition and serial position. A main effect of talker was not obtained. A significant main effect of serial position was obtained ($F[9,981] = 334.1$, $p < .001$). A marginally significant interaction of talker and serial position was also obtained ($F[18,981] = 1.57$, $p < .06$).

In order to investigate the interaction of talker and serial position, separate two-way ANOVAs for the factors of talker condition and serial position were conducted for the primacy region (list positions 1-3), middle region (list positions 4-7), and recency region (list positions 8-10) of the serial position curve. In the primacy region of the serial position curve, a main effect of talker was obtained ($F[2,109] = 4.41$, $p < .02$). Post-hoc Newman-Keuls tests revealed that recall of items in the single-talker condition was significantly better than recall in either of the multiple-talker conditions. The two multiple-talker conditions were not significantly different from one another. Thus, multiple-talker recall performance was not affected by increased variability due to gender-related talker differences. A significant main effect of serial position was also obtained ($F[2,218] = 380.1$, $p < .001$). The interaction of talker and serial position was not significant.

In the middle region of the serial position curve, the main effect of talker was not significant. A significant main effect of serial position was obtained ($F[3,327] = 24.15$, $p < .001$). The interaction of talker and serial position was not significant. In the recency region of the serial position curve, the main effect of talker was not significant. A significant main effect of serial position was obtained ($F[2,218] = 427.2$, $p < .001$). The interaction of talker and serial position was not significant.

The results of Experiment 1 demonstrate that the recall of words in the primacy region of the serial position curve was significantly better for items produced by a single talker compared to items produced by multiple talkers. Thus, variability from item to item in the voice of the talker produced salient effects on the recall of early list items in a serial recall paradigm. These results suggest that the processing of multiple-voice input places greater demands on limited capacity resources in short-term memory compared to the processing of speech from a single talker. This interpretation is based on the hypothesis that primacy recall is affected by the amount of rehearsal devoted to the first few items in a list. It appears that subjects in the single-talker condition obtained better primacy recall performance because early list items received more rehearsal, or more efficient rehearsal, than early list items in the multiple-talker conditions.

Why would rehearsal of items in the single-talker condition be more efficient than the rehearsal of items in the multiple-talker conditions? One explanation is that the perception of speech from multiple talkers requires more processing resources for the encoding of these list items, compared to the encoding of single-talker list items. As a result, fewer processing resources are available for the rehearsal of items from multiple-talker lists, leading to differences in primacy recall (Kahneman, 1973). Thus, differences in the efficiency or amount of rehearsal for multiple-talker and single-talker items may reflect differential capacity demands for the initial encoding of voice-specific acoustic-phonetic information in these stimuli.

Another possibility is that changes from stimulus to stimulus in the voice of the talker do not affect the speed or efficiency of initial encoding processes, but instead affect the efficiency of rehearsal processes after stimulus items have been encoded. In this case, relatively more processing resources would be needed for the rehearsal of multiple-talker items after they have been encoded. Because talker voice information varies from item to item in a multiple-talker list, listeners may not be able to extract enough talker-specific invariant cues to support efficient rehearsal processes for both item and order information. This explanation, however, is not consistent with the data reported by Mullennix et al. (1987), demonstrating that spoken word recognition is slower and less accurate when the voice of the talker changes from trial to trial. These results suggest that variability due to the voice of the talker adversely affects the speed and/or efficiency of encoding for spoken words.

The present results cannot distinguish between the hypotheses that primacy recall differences between single-talker and multiple-talker lists are due to differences in encoding and rehearsal, or just rehearsal. However, the data do suggest that the processing of multiple-voice input requires a greater amount of the limited-capacity resources in short-term memory compared to the processing of speech produced by a single talker. Compensation for talker variation does not appear to be automatic or capacity-free. There is some cost associated with changes from item to item in the voice of the talker, as measured by recall performance in this task.

Serial recall of a list of items requires subjects to encode and rehearse not only item information but also order information associated with each item. Compared to a free recall task, serial recall requires more processing resources in short-term memory and may be more likely to reveal differences in the initial encoding and/or rehearsal of single-talker and multiple-talker word items. In order to examine whether the encoding of order information is needed to produce recall differences for multiple-talker and single-talker word lists, a free recall experiment was conducted. As in Experiment 1, subjects heard word lists produced by a single talker or by different male and female talkers. In contrast to Experiment 1, subjects were free to recall the words in any order.

Experiment 2

Method

Subjects. Subjects were 40 undergraduate students at Indiana University who participated to fulfill a course requirement in introductory psychology. Each subject participated in one hour-long session. All subjects were native speakers of English who reported no history of a speech or hearing disorder at the time of testing.

Stimuli. The stimuli consisted of 15 lists of 20 monosyllabic English words spoken by a single talker or by different male and female talkers. As in Experiment 1, words were originally recorded in isolation on audiotape, digitized via a 12-bit analog-to-digital converter, and digitally equated for overall RMS amplitude. Stimuli were low-pass filtered at 4.8 kHz and played to subjects through a 12-bit digital-to-analog converter over matched and calibrated TDH-39 headphones at 80 dB SPL. Words within a list were semantically unrelated and differed from each other by at least two phonemes.

Procedure. Subjects were tested in groups of six or less in a sound-treated room. On each trial, subjects were presented with a spoken list of 20 words. They were then given 60 seconds to recall the words in any order. Subjects recorded their responses by printing them on a response sheet. The inter-word interval for stimulus presentation and the placement of warning tones at the beginning and end of each list were the same as in Experiment 1.

The talker variable was manipulated in a between-subjects design. Subjects were randomly assigned to one of two talker conditions: single-talker, in which all list items were spoken by a single talker, and multiple-talker, in which the 20 items within each list were spoken by 10 different male talkers and 10 different female talkers. Identical word lists were used in both talker conditions; the conditions differed only in terms of the talkers used to produce the words in each list. Each subject heard 15 unrelated 20-item word lists. The order of stimuli within each list was randomized.

Results and Discussion

Figure 2 shows the percentage of words correctly recalled as a function of serial position for both talker conditions.

Insert Figure 2 about here

Inspection of Figure 2 indicates that the serial position curve was obtained for both talker conditions. The primacy effect, however, does not appear to be as large as the one observed in Experiment 1. Because 10-item word lists were used in Experiment 1 and 20-item word lists were used in Experiment 2, this result is consistent with previous data demonstrating decreased primacy recall when longer lists of items are presented for recall (Murdock, 1962). In order to test for overall recall differences between talker conditions, a two-way ANOVA was conducted on the recall data examining the effects of talker condition and serial position on free recall performance. The main effect of talker was not significant. A significant main effect of serial position was obtained ($F[19,722] = 76.4$, $p < .001$). The interaction of talker and serial position was also significant ($F[19,722] = 1.67$, $p < .04$).

In order to investigate the interaction of talker and serial position, separate two-way ANOVAs were conducted for the primacy region (list positions 1-6), middle region (list positions 7-12), and recency region (list positions 13-20) of the serial position curve. In the primacy region of the serial position curve, neither the main effect of talker nor the interaction of talker and position reached significance. A significant main effect of serial position was obtained ($F[5, 190] = 21.37$, $p < .001$). Similarly, in the middle region of the serial position curve, the main effect of talker and the interaction of talker and serial position were not significant. The main effect of serial position was also not significant.

In the recency region of the serial position curve, a significant main effect of serial position was obtained ($F[7,266] = 126.15$, $p < .001$). Although the main effect of talker was not significant, the interaction of talker and serial position was significant ($F[7,266] = 3.25$, $p < .01$). In order to test the interaction of talker and serial position, post-hoc

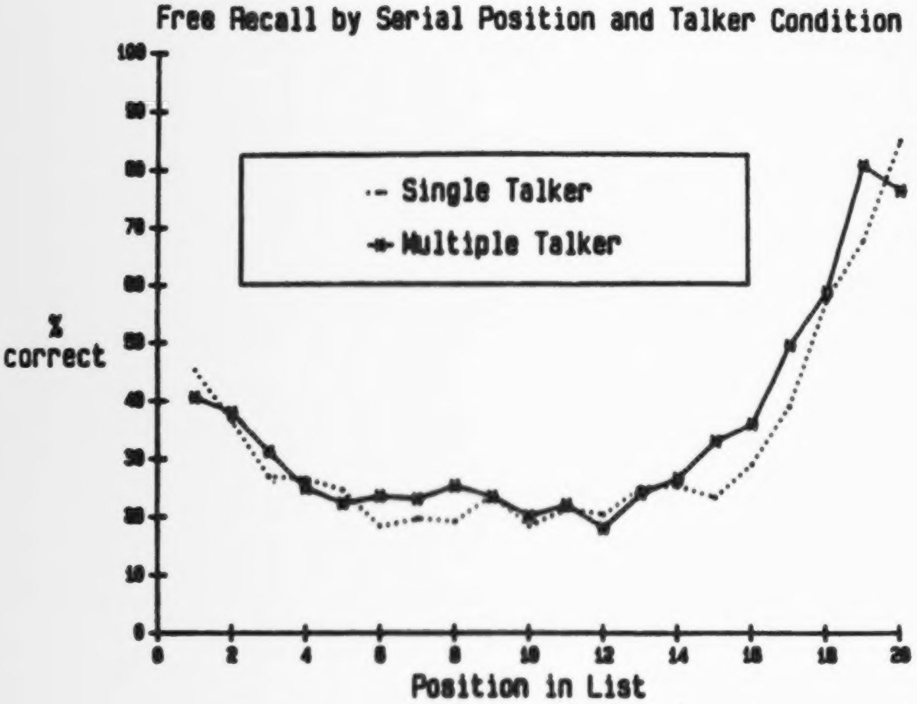


Figure 2. Mean percent correct free recall collapsed over subjects as a function of serial position and talker condition for Experiment 2.

Newman-Keuls tests were conducted comparing the recall performance of the two talker groups at each list position from 13-20. The results of these tests revealed that multiple-talker recall was better than single-talker recall at list positions 15, 17, and 19. Single-talker recall was better than multiple-talker recall at list position 20.

In summary, no consistent differences in recall performance were obtained between the multiple-talker and single-talker conditions in the free recall task. This pattern of results is in marked contrast to the superior primacy recall performance of the single-talker condition in the serial recall task in Experiment 1. Apparently, requiring subjects to encode order information in a serial recall task is an important factor in obtaining differences between single-talker and multiple-talker conditions in the recall of early list items. Experiments 1 and 2 differed in the number of list items presented to subjects, and this may have had an effect on the results obtained in the two experiments. It is not clear, however, why an increase in list length would reduce differences between the talker conditions in the recall of early list items.

Thus, it appears that the processing of multiple-voice input does not have a consistent effect on primacy recall performance unless capacity demands are increased by requiring subjects to encode both order and item information. It is possible that subjects may encode voice cues along with item and order information in serial recall. If voice cues remain the same for each item, it may be easier to associate item and order information. It may therefore be less likely that item and order information are both recalled correctly when voice cues change from item to item in serial recall.

The two talker conditions differed in the recall of items from several positions in the recency region of the serial position curve. Multiple-talker recall was better than single-talker recall at list positions 15, 17, and 19. Because the multiple-talker items contained more acoustic variability than single-talker items, a set of multiple-talker items may be more distinctive and discriminable in short-term memory than a set of single-talker items. If items differ in voice-specific acoustic information, this may contribute to their distinctiveness in short-term memory, and may facilitate maintenance rehearsal and subsequent recall of list items from the recency region of the serial position curve. However, this pattern of superior recency recall for multiple-talker lists was not observed for serial recall in Experiment 1.

Experiment 3

In order to further investigate whether the processing of items from different talkers requires greater processing resources than the processing of items from a single talker, a third experiment was conducted. In this experiment, capacity demands in short-term memory were increased by including a preload memory task along with the serial recall task (Baddeley & Hitch, 1974). The increased processing capacity required by the memory preload task should result in fewer available resources for the primary memory task (Posner & Rossman, 1965). In Experiment 3, a series of digits was visually presented on a CRT display prior to the auditory presentation of each word list. Subjects were required to recall the visually presented digits and then recall items from the spoken word list. We predicted that, as the processing resources required by the memory preload task increased, performance on both digit recall and primacy-region word recall would decrease to a greater extent for the multiple-talker lists compared to the single-talker lists.

Method

Subjects. Subjects were 72 volunteers from the Bloomington, Indiana community. Subjects participated in one hour-long session and were paid $4.00 for their participation. All subjects were native speakers of English and reported no history of a speech or hearing disorder at the time of testing.

Stimuli. The stimuli used in Experiment 1 were used in Experiment 3. All aspects of the stimuli remained the same.

Procedure. Subjects were tested in groups of six or less in a sound-treated room. The experimental procedure was identical to that used in Experiment 1, with the exception that the memory preload task was included. Prior to the auditory presentation of each word list, subjects saw either zero, three, or six digits presented sequentially on a CRT monitor directly in front of them. Each digit was sampled without replacement from the digits one through nine on each trial. Each digit remained on the CRT screen for two seconds, with a one second inter-digit interval. The placement of warning tones was the same as in Experiment 1, except that an additional tone was added to alert subjects to the beginning of the digit presentation.

Subjects were instructed to recall the visually presented preload digits and then the word list items in the exact serial order in which they were presented. During the recall interval, subjects were required to first recall the digits and then recall as many of the spoken words as possible. In order to ensure that subjects maintained the preload digits in memory during the word list presentation, they were explicitly told that none of the word items would be counted as correct unless all of the digits were correctly recalled in the exact temporal order in which they were presented.

The talker variable was manipulated in a between subjects design. Subjects were randomly assigned to one of two talker conditions: single-talker, in which all list items were spoken by the same talker, or multiple-talker, in which list items were spoken by five male and five female talkers. Memory preload was also manipulated between subjects. Subjects were randomly assigned to one of three preload conditions: no preload, three-digit preload, or six-digit preload.

Results and Discussion

Word recall and digit recall were examined separately as dependent variables. The presentation and discussion of the data is divided into two parts for ease of exposition.

Digit Recall

Because digits were not presented in the 0-digit preload condition, the analysis of the digit recall data involved only the 3-digit and 6-digit preload conditions. Digits were scored as correct if and only if they were recalled in the exact serial order in which they were presented. The percentage of digits correctly recalled as a function of talker condition and preload condition is shown in Figure 3.

Insert Figure 3 about here

A two-way ANOVA was conducted on the digit recall data for the factors of talker condition and preload condition. The analysis revealed a significant main effect of talker on digit recall (F[1,44] = 4.91, $p < .03$). Subjects in the single-talker conditions recalled 85.7% of the digits correctly while subjects in the multiple-talker conditions recalled only 78.4% of the digits correctly. A significant main effect of preload condition was also obtained (F[1,44] = 8.49, $p < .01$). A higher percentage of digits was recalled in the 3-digit preload condition (86.7%) compared to the 6-digit preload condition (77.1%).

In the three-digit preload condition, subjects in the single-talker group recalled 5.1% more digits than subjects in the multiple-talker group. This difference increased to 9.4% in the six-digit preload condition. Thus, the effect of talker variability on digit recall performance became greater in the six-digit preload condition compared to the three-digit preload condition. However, the interaction between talker and preload condition was not statistically significant (F[1,44] = 0.44, $p > .4$).

In summary, the analysis of the digit recall data from the memory preload task demonstrated that subjects recalled more digits when digit presentation was followed by a word list spoken by a single talker than a word list spoken by multiple talkers. In addition, there was a trend suggesting that differences in digit recall between the talker conditions became larger as the number of preload items was increased. These results suggest that more processing resources are required for the encoding and rehearsal of list items spoken by different talkers (Rabbitt, 1968). The digit recall data suggest that the perception of speech from multiple talkers, compared to the perception of speech from a single talker, interferes with subjects' ability to maintain information in short-term memory. The encoding and rehearsal of word lists produced by multiple talkers appears to require a greater allocation of processing resources in short-term memory.

Word Recall

The percentage of words correctly recalled as a function of talker condition and serial position is shown in the panels of Figure 4 for the 0-digit, 3-digit, and 6-digit preload conditions.

Insert Figure 4 about here

A three-way ANOVA was performed on the word recall data for the factors of talker condition, preload condition, and serial position. No significant main effect for talker was obtained. A significant main effect for preload was obtained (F[2,65] = 23.4, $p < .001$). Fewer words were recalled overall as memory preload increased. This result demonstrates that an increase in memory preload had a detrimental effect on word recall performance, suggesting that

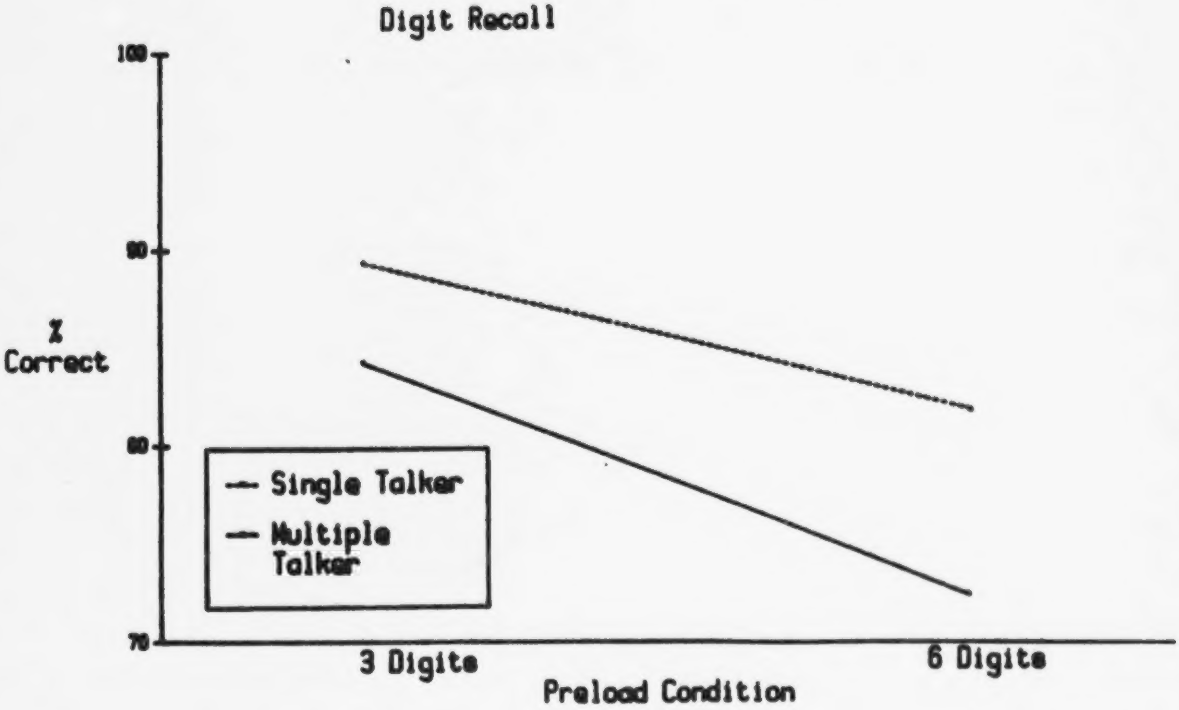


Figure 3. Mean percent correct digit recall collapsed over subjects as a function of talker condition and preload condition for Experiment 3.

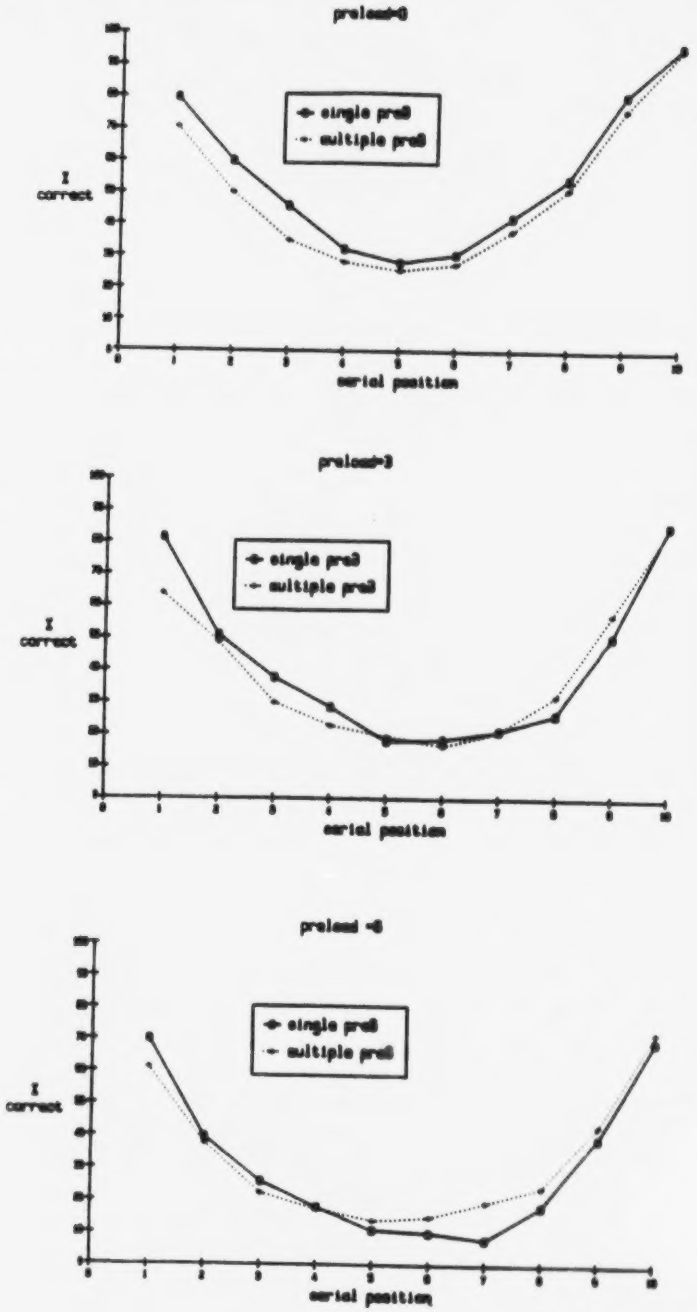


Figure 4. Mean percent correct word recall collapsed over subjects as a function of serial position and talker condition for Experiment 3. The top panel shows the 0-digit preload condition, the middle panel shows the 3-digit preload condition and the bottom panel shows the 6-digit preload condition.

the processes involved in digit and word recall share limited-capacity resources. A significant main effect of serial position was also obtained (F[9,585] = 180.56, p < .001). The interaction of talker and preload condition was not significant. A significant interaction of talker and serial position was obtained (F[9,585] = 2.0, p < .04). Examination of Figure 4 suggests that recall performance for the single-talker condition was superior only in the primacy region of the serial position curve. A significant interaction of preload condition and serial position was also obtained (F[18,585] = 2.86, p < .01). Finally, the three-way interaction was not significant.

In order to investigate the interaction of talker and serial position and the interaction of preload condition and serial position, separate three-way ANOVAs were conducted on the word recall data for the primacy region (list positions 1-3), middle region (list positions 4-7), and recency region (list positions 8-10) of the serial position curve. These analyses were conducted for the factors of talker condition, preload condition, and serial position. For recall performance in the primacy region of the serial position curve, a marginally significant main effect of talker was obtained (F[1,65] = 3.9, p < .06). Better recall was observed in the single-talker condition compared to the multiple-talker condition. A significant main effect of serial position was obtained (F[2,130] = 242.7, p < .001). A significant main effect of preload condition was also obtained (F[2,65] = 4.37, p < .02). Newman-Keuls post-hoc tests revealed that as the number of preload items increased, primacy recall decreased. However, the interaction of talker and preload condition was not significant. This result demonstrates that the preload manipulation did not reliably affect the differences between single-talker and multiple-talker primacy recall. No other interactions approached significance in this analysis. In summary, the recall of early list items was better for the single-talker condition compared to the multiple-talker condition. Recall of early list items decreased as the number of preload items increased. However, the preload manipulation did not affect the differences between single-talker and multiple-talker recall in the primacy region of the serial position curve.

In the middle region of the serial position curve, a significant main effect of talker was not obtained. A significant main effect of serial position was obtained (F[3,195] = 4.96, p < .01). A significant main effect of preload condition was also obtained (F[2,65] = 11.22, p < .001). Newman-Keuls post-hoc tests revealed that as the number of preload digits increased, word recall decreased. No significant interactions were obtained. In summary, the word recall results for the middle region of the serial position curve revealed no differences between single-talker and multple-talker word recall. As the number of preload items increased, word recall decreased.

In the recency region of the serial position curve, a main effect of talker was not observed. A significant main effect of serial position was obtained (F[2,130] = 376.9, p < .001), and a significant main effect of preload condition was also obtained (F[2,65] = 30.47, p < .001). Newman-Keuls post-hoc tests revealed that as the number of preload items increased, word recall decreased. No significant interactions were obtained. The analysis of word recall in the recency region of the serial position curve revealed no differences between the single-talker and multiple-talker conditions. In addition, recall performance decreased as the number of preload items increased.

We predicted that as the number of preload memory items increased, performance on both digit recall and primacy-region word recall would decrease, and that these effects would be greater for the multiple-talker condition than the single-talker condition. This prediction was not supported. Although a trend was observed in the digit recall data suggesting that differences between the talker conditions became larger as preload increased, this interaction was not statistically significant. In addition, differences in primacy word recall between the talker conditions did not become larger as preload increased.

Nevertheless, the results of Experiment 3 do provide support for the hypothesis that more processing resources are required for spoken word lists produced by multiple talkers. It is possible that the manner in which memory load was manipulated via the digit preload task prevented rehearsal differences from being reflected in primacy recall performance. In this experiment, subjects were presented with the preload digits before the presentation of the spoken word lists. Given this procedure, any differences in the processing of single-talker and multiple-talker lists are more likely to be observed for digit recall performance rather than word recall; the digits were presented first to subjects and therfore more rehearsal could be devoted to the digit items compared to the word items (see Crowder, 1976). If the digits are considered to be a part of each list, one would expect to observe larger differences between the talker conditions for the recall of the digit items compared to the other items in the "list". This pattern of results was, in fact, exactly what we found; more digits were recalled by subjects who listened to word lists spoken by a single talker than word lists spoken by multiple talkers. This pattern of results is also similar to the findings obtained by Luce, Feustel, and Pisoni (1983) in their study of the recall of natural and synthetic speech using a memory preload task. More digits were recalled by subjects who listened to lists of natural speech compared to synthetic speech, but the amount of preload did not affect the word recall differences between natural and synthetic speech.

Taken together, the digit recall and word recall data suggest that the encoding and/or rehearsal of multiple-talker lists requires a greater amount of processing resources in short-term memory compared to single-talker lists. The digit recall data provide strong evidence that the recall of preload digit items was attenuated by the subsequent presentation of word lists produced by different talkers. Multiple-talker word recall was not significantly greater than single-talker word recall in any region of the serial position curve for any of the three preload conditions. The superiority of single-talker digit recall appears to be due to differences in processing capacity required for single-talker and multiple-talker word lists. In summary, the results of Experiment 3 suggest that changes from item to item in the voice of the talker require more processing resources in short-term memory. The increased demands on processing resources for lists that vary from item to item in the voice of the talker appear to affect rehearsal and subsequent transfer of items into long-term memory.

Experiment 4

The results of the first three experiments provide evidence that the processing of word lists produced by different talkers requires a greater allocation of processing resources compared to word lists produced by a single talker. Differences in primacy recall performance between the two talker conditions appear to be due to differences in the amount and/or efficiency of rehearsal. The differential rehearsal explanation for these results is based on the hypothesis that the amount and efficiency of rehearsal given to early list items affects primacy recall performance. The probability of recall for early list items is thought to be a function of the amount of active processing given these items (Baddeley & Hitch, 1977; Rundis & Atkinson, 1970; Rundis, 1971).

Primacy recall performance, however, can be affected by variables other than rehearsal processes. It is possible that the differences obtained between single-talker and multiple-talker conditions in primacy recall reflect differences in search and retrieval processes that are independent of rehearsal processes. There is some evidence that a representation of talker voice characteristics can be retained in memory and used to facilitate the retrieval of words in a recognition memory task (Craik and Kirsner, 1974). If talker voice cues can be transferred into long-term memory along with associated item and order information, the redundancy of talker cues in a single-talker condition may facilitate search and retrieval processes.

One way in which retrieval processes could differentially affect the recall of spoken word lists involves the use of voice-specific cues available in short-term memory. In immediate recall paradigms, voice-specific acoustic information from terminal list items is available in short-term memory and may be used to facilitate the search and retrieval of early list items in long-term memory. If voice-specific information can be used to search long-term memory for list items, memory search may be more effective when the voice characteristics of one talker, rather than several talkers, are used during memory search. Alternatively, retrieval processes may be more effective for single-talker word lists because a set of these items are more highly associated in long-term memory compared to a set of multiple-talker word items. In this case, the previously observed primacy recall differences would be due to differences in the strength of associations among a set of items that are produced by a single talker compared to the same items produced by different talkers.

Experiment 4 was designed to assess recall performance for single-talker and multiple-talker lists when cues in short-term memory are eliminated and are not available to facilitate recall. If the differences in the recall of early list items are due to a facilitation of retrieval when the voice cues of a single talker in short-term memory are used in search, then differences in primacy-region recall between the talker conditions should not be obtained when the contents of short-term memory are eliminated by an interference task. If the previously obtained primacy recall differences are due to differences in the strength of associations among items, then single-talker recall should be greater than multiple-talker recall across all list positions when subjects must rely exclusively on long-term memory for the recall of all list items.

Experiment 4 employed a retroactive interference task with serial recall (Peterson & Peterson, 1959). Subjects were presented with a list of spoken words for serial recall and then performed an arithmetic task designed to eliminate the rehearsal of items in short-term memory before the recall period. The use of the arithmetic task is designed to occupy short-term memory, forcing subjects to rely on long-term memory for the recall of list items. The retroactive interference task should eliminate any contribution of voice-specific acoustic cues in short-term memory for retrieval of early list items. Thus, any differences between the talker conditions in the primacy region of the serial position curve should reflect differences in the amount or efficiency of rehearsal processes used to transfer items into long-term memory. Recall performance in this paradigm should not reflect differences in the cues available in short-term memory at the time of recall.

Method

Subjects. Subjects were 108 undergraduates at Indiana University who volunteered to fulfill a course requirement. Each subject participated in one hour-long session. All subjects were native speakers of English and reported no history of a speech or hearing disorder at the time of testing.

Stimuli. The stimuli used in Experiments 1 and 3 were also used in Experiment 4. All aspects of the stimuli remained exactly the same.

Procedure. The experimental procedure was identical to Experiment 1, except that a retroactive interference task was included at the end of each list. After the presentation of each spoken word list, subjects saw a three-digit number presented visually on a CRT monitor. The three digits in each number were randomly sampled without replacement from the digits one through nine and were presented simultaneously on the CRT monitor. Subjects were required to silently count backwards by three's from this three-digit number, subtracting three every time they heard a signal tone over their headphones. These tones occured at two-second intervals after the presentation of the three-digit number. The end of the arithmetic task was signaled by the presentation of two sequential tones.

After subjects heard the two tones, they were required to write down the number they currently had in memory for the subtraction task. After writing down this number, subjects were instructed to recall the items presented in the word list by writing down their responses on a response sheet. Subjects were told that their recall of word list items would be counted as correct only if they were in the correct serial position. In order to ensure that subjects paid full attention to the arithmetic task, they were told that their recall responses for the word lists would not be scored unless they produced the correct number from the subtraction task at the beginning of the recall period. Talker variability was manipulated in a between subjects design. Subjects were randomly assigned to one of two talker conditions: single-talker or multiple-talker. The length of the retroactive interference interval was also manipulated between subjects to produce three conditions: four seconds, eight seconds, and 12 seconds.

Results and Discussion

Figure 5 shows the percentage of words correctly recalled as a function of talker condition and serial position in panels for the four-second, eight-second, and 12-second retroactive interference conditions. In addition, the data obtained in Experiment 1 are replotted in the top panel as a zero-second interference condition. The data from experiment 1 were used in statistical analysis as a 0-second interference (immediate recall) control condition.

Insert Figure 5 about here

Inspection of Figure 5 reveals that the recall of items in the primacy region of the serial position curve is consistently higher for the single-talker condition compared to the multiple-talker condition for all levels of the interference variable. Recall performance in the middle and recency portions of the curve does not appear to differ between talker conditions. A three-way ANOVA was conducted on the recall data to confirm these observations. Three factors were entered into the analysis: talker condition, serial position, and duration of the interference interval.

A significant main effect of talker was obtained ($F[1,128] = 14.7$, $p < .001$). Overall percent correct recall in the single-talker condition was better than recall in the multiple-talker condition. A significant main effect of interference condition was also obtained ($F[3,128] = 32.5$, $p < .001$). Word recall decreased as the duration of the retroactive interference interval increased. Finally, a significant main effect of serial position was also obtained ($F[9,1152] = 333.3$, $p < .001$).

The interaction of talker and interference interval was not significant. Thus, differences between the talker conditions did not change as a function of interference condition. As expected, a significant interaction of serial position with interference interval was obtained ($F[27,1152] = 12.6$, $p < .001$). Recall of items from the last few serial positions decreased to a greater degree than recall of items from the other serial positions as the duration of the interference interval increased. A significant interaction of talker and serial position was also obtained ($F[9,1152] = 11.1$, $p < .001$). The three-way interaction was not significant. In summary, the analysis over all list positions revealed that recall in the single-talker conditions was better overall than recall in the multiple-talker conditions and recall differences between the talker conditions did not change as the duration of the interference interval increased. In addition, the retroactive interference task reduced recall for items in the recency region of the serial position curve to a greater degree than other items.

In order to investigate the interaction of talker and serial position, and the interaction of interference condition and serial position, separate three-way ANOVAs were carried out for the primacy region (list positions 1-3), middle region (list positions 4-7), and recency region (list positions 8-10) of the serial position curve. For the primacy region of the serial position curve, a significant main effect of talker was obtained ($F[1,128] = 52.9$, $p < .001$). Recall of items from early list positions was greater for the single-talker condition than the multiple-talker condition. The main effect of interference interval was not significant. A significant main effect of serial position was obtained ($F[2,256] = 474.4$, $p < .001$). No other significant interactions were obtained. In summary, the analysis of recall performance for the primacy region of the serial position curve revealed that recall in the single-talker condition was greater than recall in the multiple-talker condition. This difference did not change significantly as a function of interference condition.

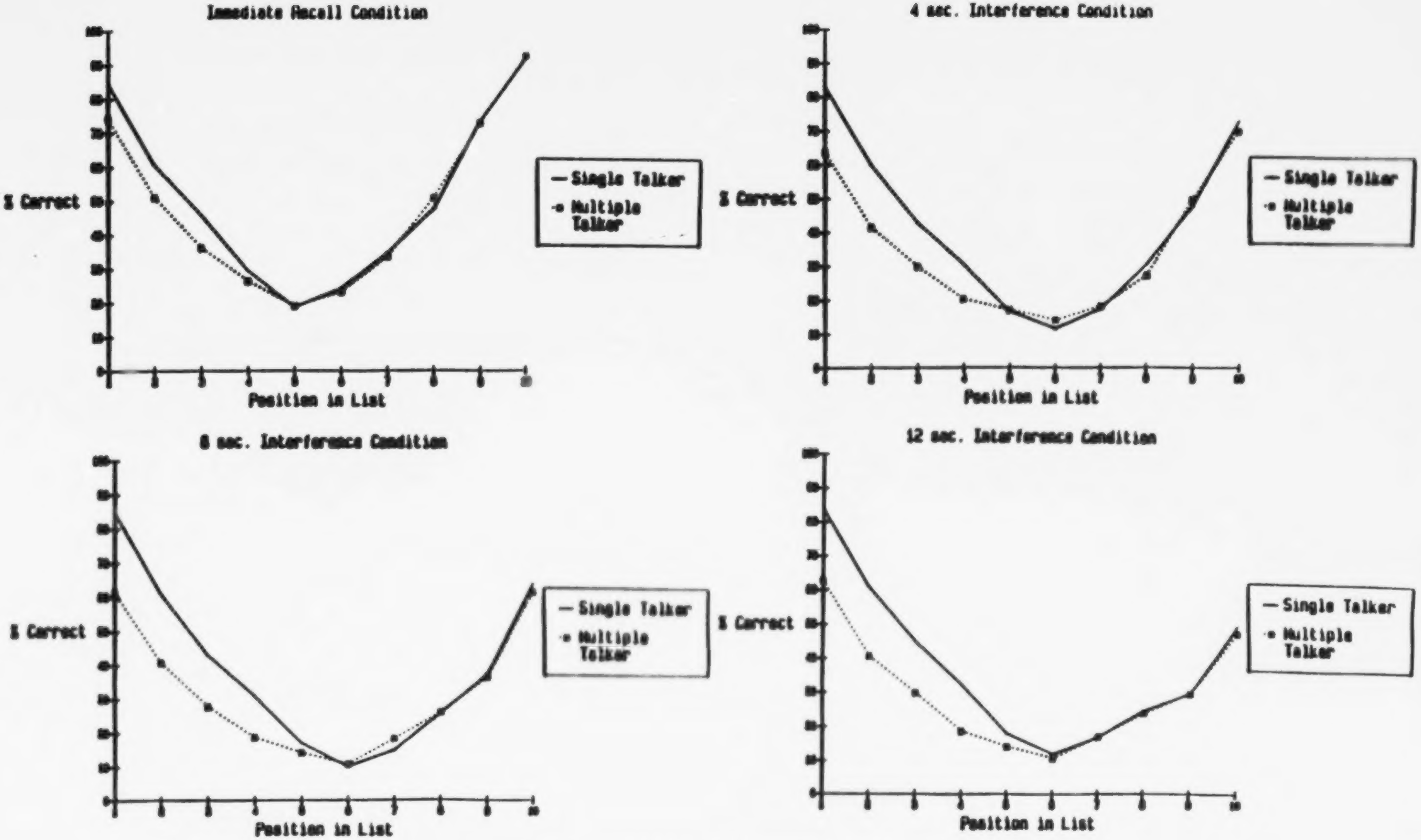


Figure 5. Mean percent correct recall collapsed over subjects as a function of serial position and talker condition for Experiment 4. Panel A shows recall data from the immediate recall condition in Experiment 1; panel B shows recall data from the 4-second interference condition; panel C shows recall data from the 8-second interference condition; and panel D shows recall data from the 12-second recall condition.

For the middle region of the serial position curve (list positions 4-7), the main effect of talker was not significant. A significant main effect of interference condition was observed ($F[3,128] = 8.26$, $p < .001$). Recall performance became worse as the duration of the interference interval increased. A significant main effect of serial position was also observed ($F[3,384] = 25.1$, $p < .001$). The interaction of talker and interference condition was not significant. A significant interaction of talker and serial position was also obtained ($F[3,384] = 6.14$, $p < .001$). Newman-Keuls post-hoc tests revealed that recall was better for the single-talker condition than the multiple-talker condition at serial position 4, but that the talker conditions did not differ at serial positions 5, 6, and 7. The superior recall for the fourth list position by the single-talker group is consistent with the pattern of results in the primacy region of the serial position curve, as this position could be considered part of the primacy region of the serial position curve.

For the recency region of the serial position curve (list positions 8-10), a main effect of talker was not obtained. A significant main effect of serial position was obtained ($F[2,256] = 330.5$, $p < .001$) along with a significant main effect of interference interval ($F[3,128] = 70.8$, $p < .001$). Newman-Keuls post-hoc tests revealed that recall in the immediate recall condition was better than recall in the 4, 8, and 12 second interference conditions, and that recall in the 4-second interference condition was better than recall in the 8 and 12 second interference conditions. Recall in the 8 and 12 second conditions did not differ reliably. No significant interactions were obtained. These results, taken together with the absence of a main effect for talker condition, suggest that recall in the recency region of the serial position curve did not differ for the single-talker and multiple-talker groups in any of the interference conditions.

The results of Experiment 4 revealed that variability in the voice of the talker produced effects on recall that were restricted to the primacy region of the serial position curve. As observed in Experiments 1 and 3, single-talker recall was superior to multiple-talker recall for early list items. No differences in recall performance were observed for the single-talker and multiple-talker conditions in the middle and recency regions of the serial position curve at any duration of the interference task.

The interference task was designed to occupy short-term memory, thereby forcing subjects to rely on long-term memory for the recall of list items. To the extent that the interference task eliminated the contents of short-term memory, recall performance did not reflect any contributions of voice-specific acoustic cues in short-term memory to the retrieval of list items. The recall differences in the primacy region of the serial position curve due to talker variability were not related to the length of the interference task interval. These results suggest that recall of items from the primacy region of the serial position curve is independent of processes operating on the contents of short-term memory. Thus, the superior primacy recall performance of the single talker condition observed in Experiment 4 does not appear to be due to differences in voice cues available in short-term memory at the time of recall, since these cues were eliminated by the interference task.

In addition, recall performance for the talker conditions did not differ over all serial positions when subjects were forced to rely on long-term memory for recall. If the primacy recall differences obtained in the previous experiments were simply due to differences in the strength of associations among a set of items, single-talker recall should have been better than multiple-talker recall over all list positions. This result was not obtained;

single-talker and multiple-talker recall differed only for early list positions. Thus, the superior primacy recall performance of the single-talker condition appears to be due to more efficient rehearsal for word lists spoken by a single talker.

General Discussion

The serial recall of early word list items and visually presented digits was better when items within a word list were spoken by a single talker than when items were spoken by different talkers. Because the recall of early list items is affected by the amount and degree of elaboration of rehearsal processes, it appears that fewer processing resources are available for the rehearsal of list items when they are produced by different talkers. Reduced primacy recall performance resulting from the perception of multiple-talker word lists are due to the increased capacity required for encoding and/or rehearsal processes which subsequently affect the transfer of items into long-term memory.

Moreover, the lack of differences in the recall of early list items between single-talker and multiple-talker conditions in the free recall experiment suggests that the increased processing capacity required for the encoding of order information in serial recall is an important factor in obtaining differences in recall performance as a function of talker variability. Apparently, increased capacity demands for the processing of multiple-voice input do not have significant effects on recall unless capacity demands are increased by procedures such as requiring subjects to encode order information. When subjects must encode and rehearse both item and order information for items that are produced by different talkers, sufficient processing resources may not be available to support efficient elaboration and transfer of items into long-term memory.

The results of the memory preload experiment provide additional support for the hypothesis that a greater amount of processing resources are required for the encoding and rehearsal of multiple-talker lists. Subjects recalled more preload digits when these digits were followed by the presentation of a single-talker, compared to a multiple-talker, word list. This result demonstrates that the processing of multiple-talker input interferes with the rehearsal and subsequent retention of digit items in memory. It appears that listeners need to allocate more processing resources when processing multiple-talker input, thereby reducing the resources available for the rehearsal of the digits.

The results of the retroactive interference experiment provide evidence that the primacy recall differences between single-talker and multiple-talker conditions are not entirely due to search and retrieval processes independent of rehearsal. Primacy recall differences were not reduced or eliminated by the retroactive interference task, suggesting that differences in the recall of single-talker and multiple-talker word lists are not due to the use of voice cues in short-term memory at the time of recall. In addition, no differences in recall between the talker conditions were obtained for the middle and recency regions of the serial position curve. Thus, differences in the recall of early list items are not simply due to stronger associations in long-term memory among a set of items produced by the same talker. If this explanation were correct, single-talker recall would have been better than multiple-talker recall across all serial positions in the list. Instead, recall differences between the talker conditions were restricted to the primacy region of the serial position curve.

Overall, the results of the present set of experiments support the hypothesis that the encoding and/or rehearsal of spoken words produced by different talkers requires a greater allocation of processing resources in short-term memory compared to items produced by a single talker. The increased processing resources required for multiple-talker lists reduces the ability of subjects to support rehearsal processes for list items. The precise nature of the rehearsal differences between single-talker and multiple-talker list items is not clear at this time. In multistore models of memory, rehearsal has been defined in terms of the number of rehearsals given an item (e.g. Atkinson & Shiffrin, 1968; Waugh & Norman, 1965). Within this framework, the processing capacity required for multiple-talker word lists reduces the number of rehearsals given to list items, thus reducing the probability of retrieval from a long-term store.

Some theorists have defined rehearsal as any active processing that keeps information available in consciousness (Dark & Loftus, 1976) and have distinguished between different types of rehearsal processes. Craik and Lockhart (1972) have described two types of rehearsal. Type I rehearsal maintains information during processing but does not lead to a more durable memory trace. Type II rehearsal involves deeper and more elaborative processing of items and leads to a more durable memory trace. Craik and Watkins (1973) called Type I rehearsal "maintenance rehearsal" and Type II rehearsal "elaborative rehearsal". According to these investigators, elaborative rehearsal serves to "enrich and elaborate" a memory trace, leading to increased retention. Maintenance rehearsal keeps items active in consciousness but does not increase the probability of retention. Within this framework, the superior recall of early list items for lists spoken by a single talker may reflect a greater amount of elaboration given to these items. Variability from item to item due to the voice of the talker may reduce the amount of elaborative rehearsal that can be given to list items.

The present results demonstrate that certain well-known experimental paradigms in memory research can be used profitably to investigate the capacity demands required for transferring speech input into memory. Our results are consistent with the hypothesis that the perceptual system utilizes some sort of talker normalization mechanism or process to encode speech produced by different talkers. Normalization for talker is not capacity free, and has consequences not only for perception, but for memory processes as well. Perceptual and memory systems appear to encode and maintain variability in stimulus input, as demonstrated by the effects of talker variability on perceptual and memory tasks.

References

Allard, F., & Henderson, L. (1976). Physical and name codes in auditory memory: the pursuit of an analogy. Quarterly Journal of Experimental Psychology, 28, 475-482.

Atkinson, R.C., & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence & J.T. Spence (Eds.), The psychology of learning and motivation (Vol. 2, pp. 89-105). New York: Academic Press.

Baddeley, A.D., & Hitch, G.J. (1974). Working Memory. In G.H. Bower (Ed.) The psychology of learning and memory (Vol. 8). New York: Academic Press.

Baddeley, A.D., & Hitch, G.J. (1977). Recency re-examined. In S. Dornic (Ed.) Attention and performance (Vol. 6, pp. 647-667). Hillsdale, N.J.: Erlbaum.

Brodie, D.A., & Prytulak, L.S. (1975). Free recall curves: nothing but rehearsing some items more or recalling them sooner? Journal of Verbal Learning and Verbal Behavior, 14, 549-563.

Bruce, D., & Crowley, J.J. (1970). Acoustic similarity effects on retrieval from secondary memory. Journal of Verbal Learning and Verbal Behavior, 9, 190-196.

Bruce, D., & Papay, J.P. (1970). Primacy effect in single-trial free recall. Journal of Verbal Learning and Verbal Behavior, 9, 473-486.

Cole, R.A., Coltheart, M., & Allard, F. (1974). Memory of a speaker's voice: reaction time to same- or different-voiced letters. Quarterly Journal of Experimental Psychology, 26, 1-7.

Craik, F.I.M., & Levy, B.A. (1970). Semantic and acoustic information in primary memory. Journal of Experimental Psychology, 86, 77-82.

Craik, F.I.M., & Lockhart, R.S. (1972). Levels of processing: a framework for memory research. Journal of Verbal Learning and Verbal Behavior, 11, 671-684.

Craik, F.I.M., & Watkins, M.J. (1973). The role of rehearsal in short-term memory. Journal of Verbal Learning and Verbal Behavior, 12, 599-607.

Craik, F.I.M. (1973). A "levels of analysis" view of memory. In P. Pliner, L. Krames, & T. Alloway (Eds.), Communication and affect: language and thought. New York: Academic Press.

Craik, F.I.M., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. Quarterly Journal of Experimental Psychology, 26, 274-284.

Creelman, C. D. (1957). Case of the unknown talker. Journal of the Acoustical Society of America, 29, 655.

Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.

Dark, V.J., & Loftus, G.R. (1976). The role of rehearsal in long-term memory performance. *Journal of Verbal Learning and Verbal Behavior*, *15*, 479-490.

Glanzer, M., & Cunitz, A.R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, *5*, 351-360.

Glanzer, M. (1972). Storage mechanisms in recall. In G.T. Bower and J.T. Spence (Eds.) *The psychology of learning and motivation* (Vol. 5, pp. 129-193). New York: Academic Press.

Greene, R.L. (1986). Sources of recency effects in free recall. *Psychological Bulletin*, *99*, 221-228.

Joos, M. A. (1948). Acoustic phonetics. *Language*, Suppl. 24, 1-136.

Luce, P.A., Feustel, T.C., & Pisoni, D.B. (1983). Capacity demands in short-term memory for synthetic and natural speech. *Human Factors*, *25*, 17-32.

Mattingly, I.G., Studdert-Kennedy, M., & Magen, H. (1983). Phonological short-term memory preserves phonetic detail. *Journal of the Acoustical Society of America*, suppl. 1, *73*, s4.

Miller, G. A., Heise, G., and Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, *41*, 329-335.

Mullennix, J.W., & Pisoni, D. B. (1986). Effects of talker uncertainty on auditory word recognition: a first report. *Research on speeech perception progress report no. 12*. Bloomington, IN., Speech Research Laboratory, Department of Psychology, Indiana University.

Mullennix, J.W., Martin, C.S., & Pisoni, D.B. (1987). Some effects of talker variability on spoken word recognition. *Research on speech perception progress report no. 13*. Bloomington, IN., Speech Research Laboratory, Department of Psychology, Indiana University.

Murdock, B. B., Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *18*, 206-211.

Peterson, L. J., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*, 193-198.

Posner, M. I., & Rossman, E. (1965). Effect of size and location of informational transforms upon short-term retention. *Journal of Experimental Psychology*, *67*, 496-505.

Rabbitt, P. (1968). Channel capacity, intelligibility, and immediate memory. *Quarterly Journal of Experimental Psychology*, *20*, 241-248.

Rundis, D., & Atkinson, R.C. (1970). Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior*, *9*, 99-105.

Rundis, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, *89*, 43-50.

Shiffrin, R.M. (1970). Memory Search. In D. Norman (Ed.), *Models of human memory* (pp.375-447). New York: Academic Press.

Shiffrin, R.M. (1976). Capacity limitations in information processing, attention, and memory. In W.K. Estes (Ed.), *Handbook of learning and cognitive processes, Vol. 4*. Hillsdale, N.J.: Erlbaum.

Summerfield, Q., & Haggard, M.P. (1975) Vocal tract normalisation as demonstrated by reaction times. *Report on research in progress in speech perception*, *2*, 1-12. The Queen's University of Belfast, Belfast, Northern Ireland.

Summerfield, Q. (1973). Acoustic and phonetic components of the influence of voice changes and identification times for CVC syllables. *Report on research in progress in speech perception*, *2*, 73-98. The Queen's University of Belfast, Belfast, Northern Ireland.

Sussman, H.M. (1986). A neuronal model of vowel normalization and representation. *Brain and Language*, *28*, 12-23.

Verbrugge. R.R., Strange, W., Shankweiler, D.P., & Edman, T.R. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America*, *60*, 198-212.

Waugh, N.C., & Norman, D.A. (1965). Primary memory. *Psychological Review*, *72*, 89-104.

[RESEARCH ON SPEECH PERCEPTION Progress Report No. 13 (1987) Indiana University]

The Perception of Digitally Coded Speech by Native and Non-native Speakers of English*

Kazunori Ozawa and John S. Logan

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, IN 47405

---

*The first author is also with C & C Information Technology Research Laboratories, NEC Corporation, Kawasaki 213, Japan. This study was carried out when the first author was a visiting scientist at the Speech Research Laboratory in Indiana University. The authors would like to thank Prof. D. B. Pisoni for his encouragement and many useful suggestions throughout this study. We also would like to thank Dr. B. G. Greene, J. Charles-Luce, and L. Huber for their efforts in recruiting non-native subjects and conducting the experiments reported in this paper, and M. J. Dedina for his help in computer programming. This research was supported, in part, by NIH grant NS-12179-11, and, in part, by NSF grant IRI 86-17847 to Indiana University.

Abstract

The segmental intelligibility of both unprocessed and coded speech was measured using the modified rhyme test (MRT). To investigate not only perceptual differences between unprocessed and coded speech, but also how language knowledge and experience affect perception, both native and non-native speakers of English served as listeners. Unprocessed speech was compared to 8 kb/s pitch predictive multi-pulse excited speech coding (MPC) and 50 kb/s u-law PCM speech (PCM). For native speakers of English, the intelligibility of unprocessed speech was the best followed by PCM and then MPC. For non-native speakers of English, the intelligibility of coded speech was much worse than unprocessed speech when compared with the results obtained from native speakers of English. The intelligibility of PCM for non-native listeners was not reliably different from MPC, although the bit rate of PCM was more than six times as high as MPC. Non-native speakers also had a tendency to confuse stop and fricative consonants, especially in coded speech, more than native speakers. These results suggest that language knowledge and experience may play a more important role in the perception of coded speech than in the perception of unprocessed speech. Further, non-native speakers may be more substantially affected by certain characteristics of the noise present in coded speech than native speakers of English. The results have implications for the design and implementation of low bit-rate speech coders.

The Perception of Digitally Coded Speech by Native and Non-native Speakers of English

Much research has been carried out on low bit rate speech waveform coding methods such as multi-pulse excited coding (Atal & Remde, 1982; Araseki, Ozawa, Ono, & Ochiai, 1983; Ozawa & Araseki, 1986) and stochastic LPC coding (Schroeder & Atal, 1985; Transco & Atal, 1986; Atal, 1987) to produce high-quality speech at bit rates below 10 kb/s. However, acoustic cues provided in reconstructed speech by the low bit rate coding methods may be impoverished compared to those found in unprocessed speech. As a consequence, language knowledge and experience may therefore become much more important for perceiving coded speech than for unprocessed speech since listeners must compensate for a lack of acoustic-phonetic redundancies by using various sources of language knowledge. Accordingly, the study of speech coding methods by not only native speakers but also non-native speakers is important in order to investigate how language knowledge and experience affects speech perception. Furthermore, from the viewpoint of developing more sensitive evaluation methods for coded speech, evaluation by non-native speakers of English using English stimuli may be very informative and useful for further improving the speech quality of coding methods. Non-native listeners may be more sensitive to small amounts of acoustic degradation caused by the coding methods than native listeners and may reveal different patterns of errors in their performance.

With regard to the effects of language knowledge and experience on speech perception, several previous studies have investigated the relationship between the perception of speech and language proficiency. Using unprocessed natural speech, Gat and Keith (1978) studied the effect of linguistic experience on auditory discrimination of words at various signal-to-noise ratios and found that word identification by non-native listeners became much poorer than native listeners when the noise level was increased. Nooteboom and Doodeman (1980) used a gating task to study the recognition of isolated words and found significant differences in gating duration at recognition points for native and non-native listeners. In our laboratory, Greene (1986) examined the relationship between the intelligibility of synthetic speech and the language proficiency of non-native listeners and found that the correlation between intelligibility of sentences for non-native listeners and their linguistic ability was high. She also suggested that synthetic speech could be used for measuring language proficiency of non-native listeners. To study the relationship between the perception of vocoded speech and language knowledge of the listeners, Mack (1987) has recently examined differences in word identification using unprocessed and vocoded semantically anomalous sentences with English monolinguals and German-English bilinguals. She found that bilinguals produced many more errors than monolinguals for both unprocessed and vocoded speech. She also suggested possible differences in perceptual strategies between monolinguals and bilinguals.

In a related area, several studies (Gaies, Gradman, & Spolsky, 1977; Spolsky, Sigurd, Sato, Walker, & Arterburn, 1968) have been carried out to develop a procedure to differentiate non-native speakers of English into various levels of proficiency. These studies have shown that speech presented under various noise conditions maybe be useful for evaluating English proficiency in non-native speakers. However, these studies make use of a known amount of signal degradation (i.e., the signal-to-noise ratio) to differentiate levels of English proficiency. In the present experiment, we were interested in doing the converse: We wanted to use a group of subjects that we assumed would be less proficient with English, that is, non-native speakers of English, to differentiate various types of coded speech and unprocessed speech. Little research has been carried out on the effects of language knowledge and experience in the perception of coded speech produced by the low bit rate coding methods, and on the perceptual differences between unprocessed and coded speech using native and non-native listeners.

Several perceptual evaluation methods have been developed for assessing the perceptual quality and comprehension of speech based on the knowledge of human speech perception processes (Pisoni, 1978; Pisoni, Nusbaum, & Greene, 1985; Pisoni & Luce, 1986). Intelligibility tests have been widely used as a measure for assessing speech quality and various methods of evaluating intelligibility have been developed (Kalikow, Huggins, Blackman, Vishu, & Sullivan, 1976). Unfortunately, intelligibility scores are not sensitive measures of performance when comparing small differences among high quality speech systems (Nakatani & Dukes, 1973; Pisoni, Manous, & Dedina, 1986). However, intelligibility tests are generally useful when the differences between different kinds of speech are fairly large. Moreover, intelligibility scores may be extremely useful for diagnostic purposes, such as determining the reasons why some phonemes are less intelligible than others in various speech communications systems.

To measure segmental intelligibility, a number of tests have been developed including the Phonetically Balanced (PB) words (Egan 1948), the Rhyme test (Fairbanks, 1958), the Modified Rhyme Test (MRT) (House, Williams, Hecker, & Kryter, 1965), the Diagnostic Rhyme Test (DRT) (Voiers, Cohen, & Mickunas, 1965) and the Consonant Recognition Test (CRT) (Preusse, 1969). The MRT has been used to compare synthetic speech with unprocessed natural speech (Nye & Gaitenby, 1973) and to evaluate the intelligibility of LPC systems for various talkers (Kahn & Garst, 1983). The DRT has been used extensively to evaluate differences among vocoders with different parameter conditions (Wong & Markel, 1978). The CRT has been used to evaluate the influence of distortions such as bandwidth reduction, peak clipping, and amplitude quantization on intelligibility of PCM circuits (Goodman, Goodman, & Chen, 1978).

Of these methods for evaluating intelligibility, we selected the MRT using the closed response format for the present study. The reasons were as follows. First, the MRT is a reliable method. Second, the effects of learning are small. Third, the MRT can be easily administered to a group of untrained listeners. Fourth, scoring the MRT is very easy. Fifth, confusion information for both initial and final consonants can be obtained using the closed format MRT. Finally, many studies have been done in our laboratory using the MRT to assess the perceptual quality of text-to-speech synthesis systems (Greene, Manous, & Pisoni, 1984; Greene, Logan, & Pisoni, 1986; Logan, Pisoni, & Greene, 1985; Nusbaum, Dedina, & Pisoni 1984; Pisoni, 1979, 1981, 1982; Pisoni & Hunicutt, 1980; Pisoni, Nusbaum, & Greene, 1983; Yuchtman, Nusbaum, & Pisoni, 1985). In the closed format MRT, confusion information for vowels cannot be obtained. However, for vowel perception, several studies have reported that discrimination between vowels is relatively independent of listeners' linguistic experiences (Stevens, Libermann, Studdert-Kennedy, & Ohman, 1969). Thus, for the present study, we selected the closed format MRT to study consonant perception in initial and final position.

The present study was also designed to examine how language knowledge and experience affect the perception of unprocessed speech and coded speech produced by two speech coding methods. Specifically, this paper reports the results of tests measuring the segmental intelligibility of unprocessed and coded speech using the MRT. In order to investigate not only the perceptual differences between unprocessed and coded speech but also the importance of language knowledge and experience in speech perception, both native and non-native speakers of English were used as listeners. The speech coding methods used in the present study were 8 kb/s pitch-predictive multi-pulse excited speech coding (MPC) and 50 kb/s u-law PCM coding (PCM). PCM served as a standard for comparison with the MPC in the same way that the unprocessed speech served as the baseline for both types of coded speech.

Method

Subjects. Subjects consisted of two groups: (1) seventy-two native speakers of English who were undergraduate students at Indiana University enrolled in an introductory psychology course, and (2) seventy-two non-native speakers of English with various language backgrounds living in the Bloomington area. The native speakers received class credit for their participation, while non-native speakers were paid $3.50 for their participation in the experiment. Most of the non-natives were students or spouses of students enrolled at Indiana University. Table 1 shows non-native listener's language backgrounds for the three voice conditions (unprocessed speech, PCM and MPC). All subjects reported no history of a speech or hearing disorder at the time of testing.

Insert Table 1 about here

Stimuli. Six lists of 50 CVC monosyllabic words (a total of 300 words) that comprised the MRT (House, Williams, Hecker, & Kryter, 1965) were used as stimuli. These words were uttered by one male and one female talker whose native language was English. The male talker spoke a mid-western dialect, while the female talker spoke a New York dialect. The signals were band limited through a low-pass filter with a 4.8 kHz cut-off frequency, sampled at a 10 kHz sampling frequency and then digitized by a 12 bit A/D converter using a PDP-11/34 computer.

Three voice conditions were used: unprocessed speech, 8 kb/s coded speech produced by a pitch-predictive multi-pulse exited speech coding algorithm (Ozawa & Araseki, 1986) and 50 kb/s coded speech by u-law PCM. After adjusting the RMS (root mean square) level of all the stimuli to the same value, the stimuli were output using a 12 bit D/A converter at 10 kHz and recorded on audio tape using a Crown 800 series tape recorder. A 10 second synthetic vowel /a/ was recorded at the beginning of each tape to calibrate the correct playback level from session to session. The inter-stimulus interval between test words was 4 seconds.

Pitch-predictive Multi-pulse Coding Algorithm. In the pitch-predictive multi-pulse excited coding algorithm (MPC) (Singhal & Atal, 1984; Ozawa & Araseki, 1986) shown in Figure 1, the speech production process is modeled with a combination of pulses and two kinds of synthetic filters, a pitch synthetic filter (PSF) and a spectrum envelope synthetic filter (SSF). Pitch harmonic characteristics in voiced speech are represented by the PSF and vocal tract characteristics are represented by the SSF. Amplitude and locations of

Table I

Language Backgrounds for Non-native Listeners

Language	Unprocessed	PCM	MPC	Total number	%
Korean	5	2	2	9	13
Chinese	5	4	3	12	17
Japanese	1	2	1	4	6
Malay	3	0	1	4	6
Spanish	0	1	5	4	6
Finish	1	1	2	4	6
Polish	1	2	0	3	4
Others*	8	12	10	30	42
Total number	24	24	24	72	-

* Bulgarian, French, German, Italian, etc.

excitation pulses are calculated so as to minimize the perceptually weighted error between input and synthetic speech. By using this algorithm, high-quality speech can be produced in the range from 8 through 16 kb/s.

Insert Figure 1 about here

The filter coefficients of SSF were calculated by an LPC analysis method (Itakura & Saito, 1970; Makhoul, 1975; Markel & Gray, 1976). The filter coefficients of PSF were calculated by an autocorrelation method (Rabiner & Shafer, 1978). The orders of SSF and PSF were 12 and 1, respectively. The LPC analysis window was 25.6 ms and frame shift was 20 ms. The number of pulses per frame was 11 to achieve the bit rate of 8 kb/s. The analysis and bit allocation conditions are summarized in Table 2.

Insert Table 2 about here

An efficient pulse calculation algorithm (Araseki, Ozawa, Ono, & Ochiai, 1983; Ozawa, Ono, & Araseki, 1986) was used to calculate amplitudes and locations of excitation pulses. According to the algorithm, the location m of the i-th pulse is determined by searching the location which gives the absolute maximum value of g in the following equation

$$g_i = \frac{R_{nx}(m_i) - \sum_{k=1}^{i=1} g_k R_{hh}(|m_k - m_i|)}{R_{hh}(0)}, \quad 1 < m_i, m_k \leq N \qquad (1)$$

where N denotes the number of samples in which pulses are searched for. R (m) is the cross-correlation function between the perceptually weighted speech and the perceptually weighted impulse-response of the synthetic filter, and R (m) is the autocorrelation function of the weighted impulse-response. Amplitude for the i-th pulse can be calculated from Eq.(1) using the determined location m .

Procedure. Subjects were seated in a quiet sound-treated room containing six individual cubicles, each of which was equipped with a desk and a pair of high-quality headphones. Subjects read a set of instructions that described the experimental procedure. They were told that they would hear a single isolated English word on each trial of the experiment and that their task was to indicate the word they heard on the answer sheet. Subjects were told to respond on every trial and they were encouraged to guess if they were uncertain.

Subjects were provided with a closed format response form containing six response alternatives in which either initial or final consonants were the same as the word they heard. Non-native subjects were also asked to complete a language experience questionnaire in which they rated their English proficiency according to a 4 point rating scale ranging from 1 (poor) to 4

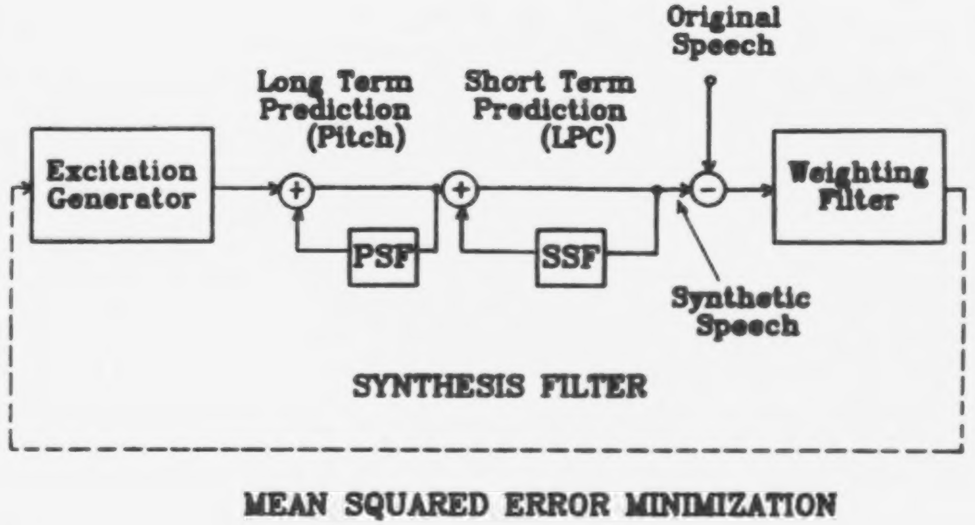

Original Speech
Long Term Prediction (Pitch)
Short Term Prediction (LPC)
Excitation Generator
+
+
−
PSF
SSF
Weighting Filter
Synthetic Speech
SYNTHESIS FILTER
MEAN SQUARED ERROR MINIMIZATION

Figure 1. Schematic diagram showing pitch-predictive multi-pulse excited speech coding algorithm.

Table II

Analysis and Bit Allocation Condition for
8kb/s Pitch Predictive Multi-pulse Excited Coding

Frame Length	20 msec
LPC Analysis Order	12
Pitch Analysis Order	1
Number of Pulses/Frame	11
Bit Allocation of LPC/Frame	45 bits
Bit Allocation of Pulses/Frame	105 bits
Bit Allocation of Pitch/Frame	10 bits

(excellent). A mixed version of the MRT was used in which items with either different initial consonants or final consonants were randomly mixed from trial to trial. Two randomizations were completed for each of the six forms of the MRT lists resulting in a total of twelve forms which corresponded to twelve experimental conditions (three voices x two talkers x two randomizations). Six subjects participated in each condition. Each subject heard all 300 MRT words. Each experimental session lasted approximately 45 minutes including instructions and a five minute break in the middle of the session.

The stimulus tapes were played back using an Ampex AG-500 tape recorder and presented binaurally over matched and calibrated Telephonics TDH-39 headphones. The signals were presented at 80 dB SPL measured by a Hewlett-Packard 400H VTVM using the calibration vowel as input. Broadband white noise (55 dB SPL) generated by a Grason-Stadler 1724 noise generator was mixed with the speech to mask the tape hiss noise.

Results and Discussion

Overall Error Rate Analysis. The data were analyzed using an analysis of variance . In the analysis of variance, listeners (native and non-native speakers of English), voices (unprocesed speech, MPC and PCM) and talkers (male and female) were between-subjects factors. Position (initial and final) was a within-subjects factor. First, the results of the analysis showed significant main effects of listeners [$F(1, 132)=157.69$, $p<.0001$], voices [$F(2, 132)=88.04$, $p<.0001$] and talkers [$F(1, 132)=60.65$, $p<.0001$]. Further, the results revealed significant interactions between listeners and position [$F(1, 132)=23.49$, $p<.0001$], voices and position [$F(2, 132)=21.48$, $p<.0001$] and talkers and position [$F(1, 132)=17.78$, $p<.0001$]. These effects are described in more detail below.

Error Rates for Native Speakers of English. Overall error rates for the native speakers for the three voice conditions averaged across talkers and consonant positions are shown in Figure 2.

Insert Figure 2 about here

The error rates were 1.3% for unprocessed speech, 3.5% for PCM and 4.6% for MPC. Post-hoc tests (Newman-Keuls) showed that the differences in error rates between each of these voices were significant. All the differences reported here may be assumed to be significant at $p<.05$.

Figure 3 shows differences in error rates between male and female talkers. For all the voices, error rates for the female talker were always slightly higher than for the male talker. Differences between male and female talkers were significant for both coding conditions but not for the unprocessed speech condition.

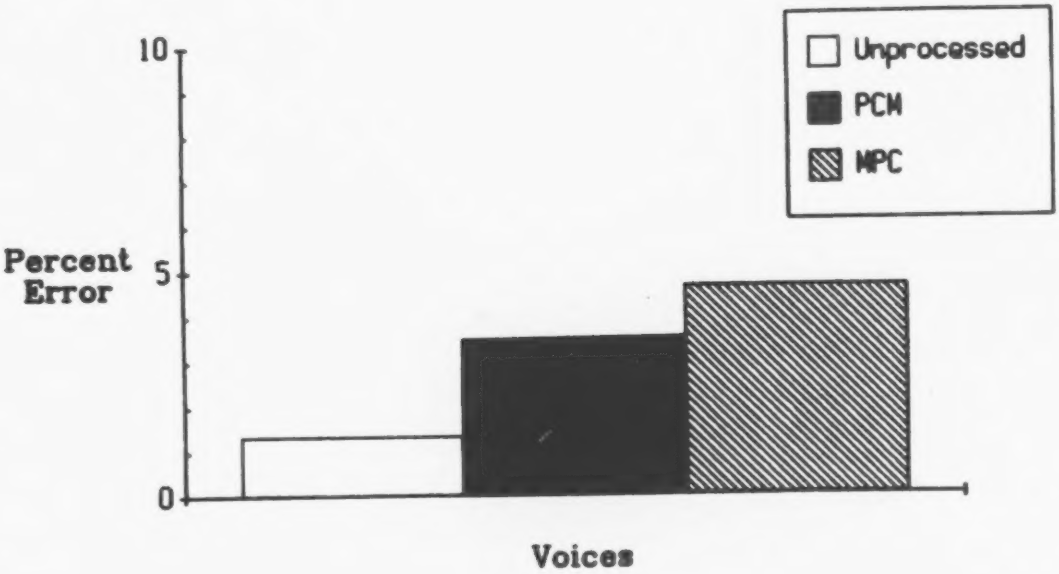


Figure 2. MRT overall error rates for the three voices (unprocessed, PCM and MPC) for native speakers of English.

Insert Figure 3 about here

Figure 4 shows differences in error rates between initial and final consonant positions. Notice that for MPC, the error rate for initial position was much higher than for final position. This difference was significant. On the other hand, differences in error rates as a function of position for the two other voice conditions were not significant. Comparing across position, differences between unprocessed and PCM speech in both initial and final position and between PCM and MPC in initial position were significant, but the difference between PCM and MPC in final position was not significant.

Insert Figure 4 about here

Consonant Confusions for Native Speakers of English. Table 3 shows the distrubution of errors as a function of manner class and consonant position. For unprocessed speech, fricatives had the highest error rates in both initial and final position. Fricatives accounted for 75% out of the 18 total errors in initial position. For PCM, fricatives were the worst in initial position, and nasals were the worst in final position. For MPC, stops were the worst in initial position and nasals were the worst in final position.

Insert Table 3 about here

Typical phoneme confusions and their error rates for the most confused manner classes in Table 3 are shown in Table 4. In initial position, the phoneme /s/ had the highest error rate and was frequently confused with the phoneme /f/ in both unprocessed speech and PCM speech. On the other hand, the phoneme /b/ had the highest error rate and was frequently confused with the phoneme /f/ in MPC.

A number of important acoustic cues for manner of articulation for stop consonants are contained in the burst and formant transitions (Borden & Harris, 1984). The burst portion and the beginning of the transition part of initial stop consonants, especially in the phoneme /b/, may be difficult to represent well in MPC because of the long duration of the analysis frame of the pulse search process as well as a lack of excitation pulses within the analysis frame. In addition, due to the error criterion in the pulse search algorithm, almost all of the excitation pulses might be used for representing the vowel part, if the burst, transition, and vowel were included in the same analysis frame. For the perception of fricatives, the frication spectra is essential (Harris, 1958). In PCM coding, the frication portion may be masked by the white noise thus contributing to the lower observed performance.

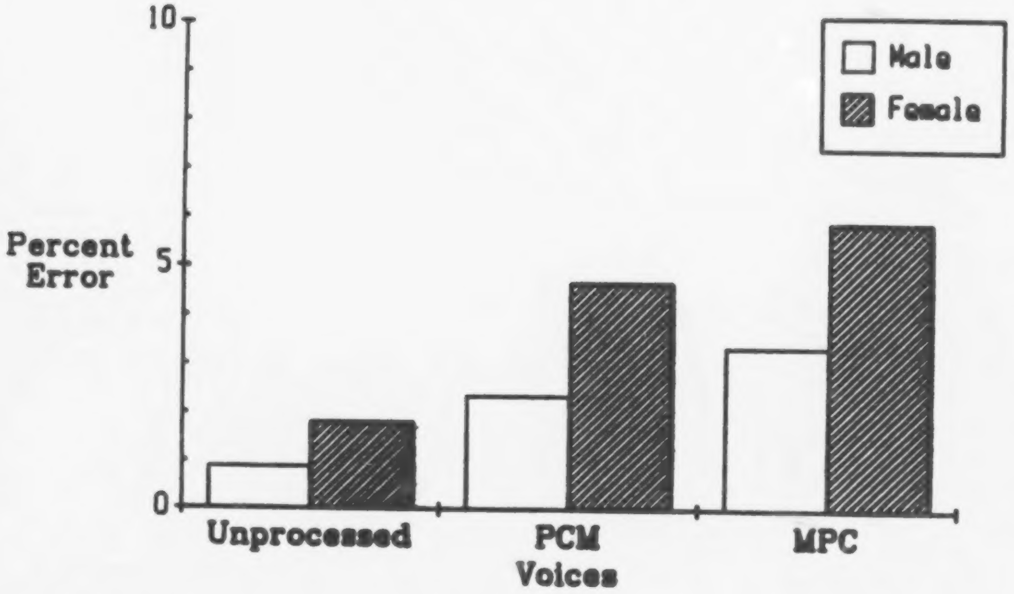


Figure 3. MRT error rates for male and female talkers for native speakers of English.

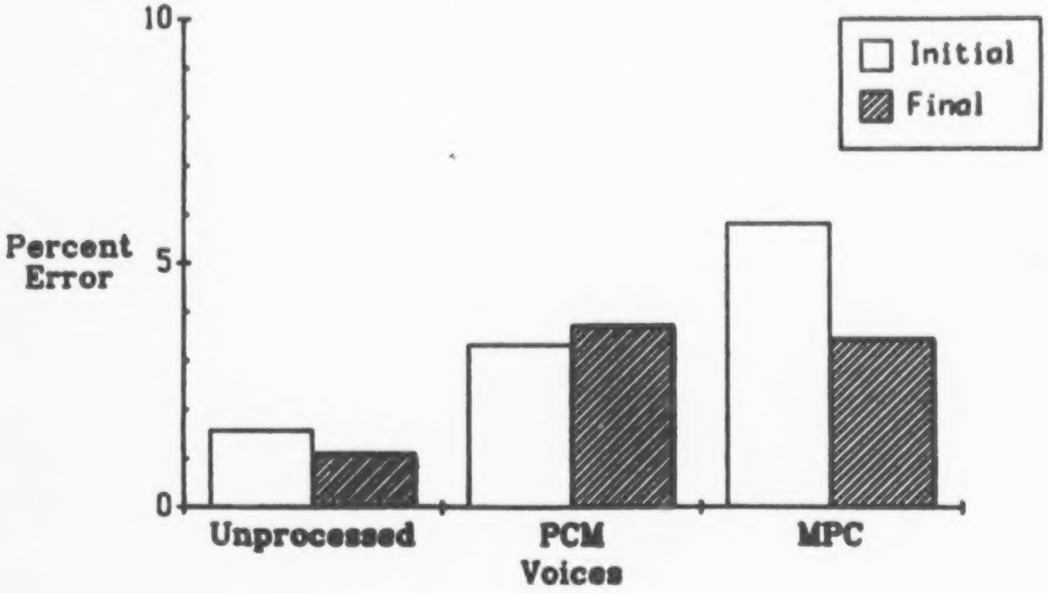


Figure 4. MRT error rates for initial and final consonant position for native and non-native speakers of English.

Table III

Errors as a Function of Manner Class for Native Speakers of English

		Initial Position		Final Position	
Voice	Manner Class	Total % of Error	Total # of Error	Total % of Error	Total # of Error
Unprocessed	Fricative	75	18	71	15
	Stop	25	6	19	4
	Nasal	0	0	10	2
PCM	Fricative	66	37	42	27
	Stop	16	9	14	9
	Nasal	18	10	44	29
MPC	Fricative	44	45	33	20
	Stop	51	53	21	13
	Nasal	5	5	46	28

For final consonants, the bilabial nasal consonant /m/ had a high error rate and was frequently confused with the alveolar nasal consonant /n/ in MPC and PCM. For final nasals, one of the important acoustic cues for the perception of place of articulation is considered to be nasalization of preceding vowels to nasal consonants (Fujimura, 1962; Hawkins & Stevens, 1985; House & Stevens, 1956; Malecot, 1960). For both MPC and PCM coding methods, such cues may not be represented adequately compared to unprocessed speech.

Insert Table 4 about here

Comparison of Error Rates between Non-native and Native Speakers of English. Figure 5 shows a comparison of the overall error rates in the three voice conditions for the non-native speakers of English and the native speakers of English. For each of three voices, non-native speakers displayed consistently higher error rates than native speakers. The differences in error rates between native and non-native speakers of English were significant for each condition. Further, differences in error rates between unprocessed speech and both types of coded speech were much higher for non-native listeners than the differences obtained from native listeners. These results suggest that language knowledge and experience may play a more important role in the perception of coded speech than in the perception of unprocessed speech. The degradation of acoustic information in coded speech appears to affect the performance of non-native speakers of English more than it affects the performance of native speakers of English who are able to compensate for the poorer quality signal by using their more extensive knowledge of English to interpret degraded or ambiguous information in the speech waveform.

Surprisingly, the difference in error rates between MPC and PCM was not significant for non-native speakers of English, although it was significant for native speakers of English. PCM contains quantization noise similar to white noise, whereas MPC has perceptually weighted noise in which the short-time spectrum envelope of the quantization noise is not white but shaped so as to reduce perceptual distortion. In this case, the short-time noise spectrum is similar to the short-time speech spectrum (Atal & Schroeder, 1979; Atal & Remde, 1982). The results obtained in both coding conditions suggest that the performance of non-native listeners may be affected by the white noise in PCM more than native listeners.

Insert Figure 5 about here

In order to examine the differences in error rates due to the amount of English language experience, the non-native subjects were divided into two groups, those with a great deal of experience with English and those with only a little experience with English. This division was carried out from analyses of the English language proficiency questionnaires given to the non-native listeners. Subjects whose rating of their experience with English in the questionnaire was greater than 3 (good) were put into the former group. The number of subjects in this group was thirty-six. Subjects whose rating of

Table IV

Typical Phoneme Confusion for the Three Voices

	Initial Position			Final Position		
	Original Phoneme	Confused Phoneme	% Error	Original Phoneme	Confused Phoneme	% Error
Unprocessed	/s/	/f/	77	/s/	/θ/	89
PCM	/s/	/f/	64	/m/	/n/	78
MPC	/b/	/f/	45	/m/	/n/	68

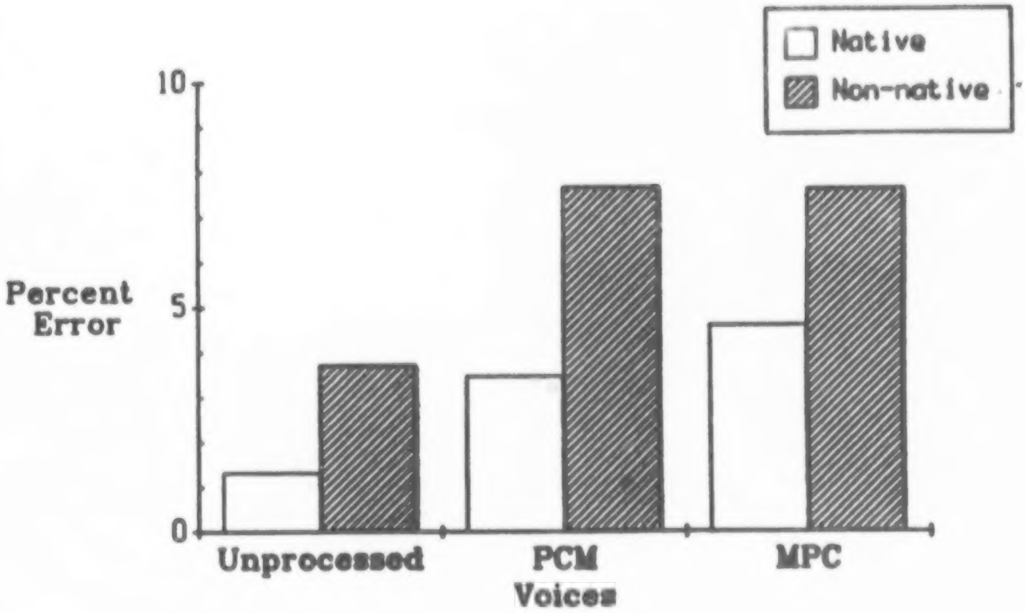


Figure 5. MRT overall error rates for the three voices (unprocessed, PCM and MPC) for native and non-native speakers of English.

their experience with English in the questionnaire was less than 2 (fair) were put into the latter group. The number of subjects in this group was also thirty-six. The results of this partitioning are shown in Figure 6. An analysis of variance showed a significant effect of experience [F(1,48)=49.68, p<.0001], and significant interactions between between position and experience [F(1,48)=8.71, p<.005], and between coding and experience [F(2,48)=3.17, p<.05]. Post-hoc tests showed that differences in error rates due to the amount of language experience were significant for PCM and MPC, but were not significant for the unprocessed speech condition. These results provide further support for the role of language knowledge and experience in speech perception, especially in the perception of coded and degraded speech (see also Greene, 1986).

Insert Figure 6 about here

Consonant Confusions for Non-native Speakers of English. Table 5 shows the distribution of perceptual errors as a function of manner class, English language experience, and consonant position obtained for the non-native speakers of English. The absolute number of errors and the proportion of the number of errors accounted for each manner class for each voice are also shown in this table.

Insert Table 5 about here

By comparing Table 5 with the data shown in Table 3, which displays the consonant confusions for the native listeners, we note the following differences. First, the number of errors in each manner class for all of the conditions was higher for the non-native listeners than for the native listeners. Second, the differences in the number of errors for manner class between native and non-native listeners was larger for consonants in final position than for consonants in initial position. The increase in the number of errors for consonants in final position was much higher for coded speech, especially for PCM as compared to unprocessed speech. Third, the percentage of errors for stop and fricative consonants was much larger for non-native listeners than for native listeners for all of the conditions. These confusions were larger for consonants in final position than for consonants in initial position, and were larger for coded speech, especially for PCM, than for unprocessed speech. Finally, for non-native listeners, the error rates for stop consonants were greater for the group of listeners with the least experience with English than for the group of listeners with the most experience with English. These findings suggest that non-native speakers of English may have a greater tendency to confuse stop and fricative consonants, especially in coded speech, than native speakers of English.

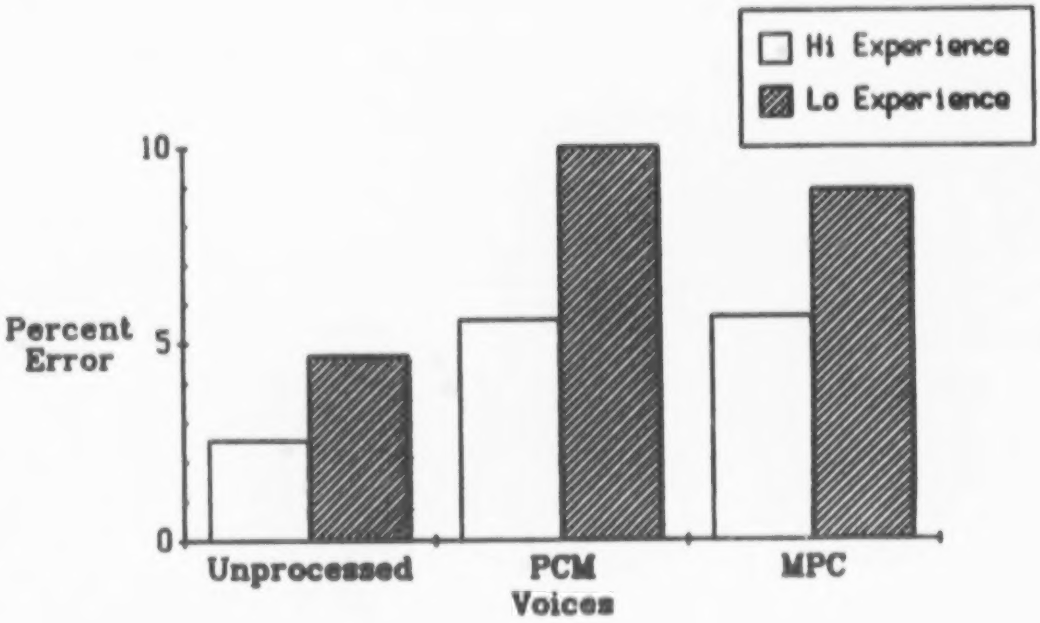


Figure 6. MRT error rates for non-native speakers of English with a large amount of experience and those with a small amount of experience with English.

Table V

Errors as a Function of Manner Class
for the Three Voices
for Non-native Speakers of English

		Initial Position # and % of Error			Final Position # and % of Error		
Voice	Manner Class	Hi Exp.	Lo Exp.	Total	Hi Exp.	Lo Exp.	Total
Unprocessed	Fricative	17 (80)	22 (73)	39 (76)	12 (52)	22 (47)	34 (49)
	Stop	4 (20)	7 (23)	11 (22)	5 (22)	18 (38)	23 (33)
	Nasal	0 (0)	1 (4)	1 (2)	6 (26)	7 (15)	13 (18)
PCM	Fricative	25 (61)	27 (61)	52 (61)	27 (39)	30 (38)	57 (34)
	Stop	11 (27)	12 (27)	23 (27)	19 (27)	23 (29)	42 (28)
	Nasal	5 (12)	5 (12)	10 (12)	24 (34)	25 (32)	49 (33)
MPC	Fricative	27 (49)	40 (48)	67 (48)	16 (43)	33 (40)	49 (41)
	Stop	24 (44)	41 (49)	65 (47)	8 (22)	29 (35)	37 (31)
	Nasal	4 (7)	3 (5)	7 (5)	13 (35)	20 (24)	33 (28)

* Percent of errors are presented in parentheses.

Hi Exp. means a large amount of experience with English and Lo Exp. means a small amount of experience with English.

General Discussion

The present study was carried out to investigate how language knowledge and experience affect the perception of unprocessed and coded speech. Our goal was to measure perceptual differences between unprocessed and coded speech for both native and non-native speakers of English. Results from the present study suggest that language knowledge and experience appears to play a much more important role in the perception of coded speech than in the perception of unprocessed speech. This result should not be surprising. In coded speech, important acoustic-phonetic information may be degraded or impoverished and the listener must compensate for the lack of acoustic-phonetic redundancies using top-down information based on various sources of language knowledge and experience. For native speakers of English, when they hear coded speech, they automatically use top-down information to compensate for the impoverished sensory information in the coded speech (Schmidt-Nielsen & Kallman, 1987). Consequently, the differences in error rates between unprocessed speech and coded speech may not be as large for native speakers of English. However, non-native speakers of English must rely on impoverished sensory information in the coded speech more than native speakers of English, since they have less resources to draw on from their knowledge and familiarity with the language. Thus, differences in error rates between unprocessed speech and coded speech would be expected to be much larger for non-native speakers of English than for native speakers of English.

The present results using coded speech also suggest that non-native speakers of English may be affected by the white noise in PCM speech much more than native speakers of English. Error analyses revealed that non-native listeners may have a tendency to confuse stop and fricative consonants more than native listeners. The confusion errors were larger for coded speech, especially in PCM, than for unprocessed speech, and they were larger for consonants in final position than for consonants in initial position.

Further study of the relationship between the specific phoneme confusion patterns associated with the perception of coded speech and specific language backgrounds of listeners should reveal not only the perceptually important acoustic cues in English but also the specific effects that language background may have on perceptual performance. In particular, the study of coded speech will have important implications for international speech communication systems using the narrow band, low bit rate speech coding methods which may be realized in the near future. If we know the perceptually important acoustic cues for specific languages, we will be able to adjust the parameters of the coding methods appropriately so as to maximize listeners' comprehension according to the specific language group to which the listeners belong. At the present time, there have been very few detailed studies of the differential effects of language background on the perception of speech sounds (see Flege, 1987).

In summary, the segmental intelligibility of unprocessed speech, 50 kb/s u-law PCM speech and 8 kb/s MPC speech was studied using the MRT. To investigate not only perceptual differences between unprocessed and coded speech but also how language knowledge and experience may affect speech perception, native and non-native speakers of English were used as listeners. For native speakers of English, the intelligibility of unprocessed speech was the best followed by PCM and then MPC. For non-native speakers of English, the differences in intelligibility between unprocessed speech and both types of coded speech were larger than that obtained with native speakers. Moreover, the difference in error rates between PCM and MPC was not significant for the non-native speakers. Non-native speakers also confused stop and fricative consonants in coded speech more than native speakers. Taken together, these results suggest that language knowledge and experience may play a much more important role in the perception of coded speech than in the perception of unprocessed speech. The results also suggest that the performance of non-native speakers of English may be more affected by the white noise in PCM than native speakers of English. The present findings suggest that speech coding methods need to be studied using both native and non-native speakers of English in order to improve speech quality under a wide variety of experimental conditions. The role of prior linguistic experience and background of the listeners has not been an important consideration in the design of efficient speech coding algorithms which are often based on speakers and listeners drawn from one uniform language population or dialect. The present findings suggest that this research strategy will need to be modified substantially in the future in order to accomodate the perceptual processing needs of non-native speakers of English who may not be able to use their knowledge of English as efficiently as native speakers normally do in a wide variety of speech communication tasks.

References

Ali, L., Gallagher, T., Goldstein, J., & Daniloff, R. (1971). Perception of coarticulated nasality. *J. Acoust. Soc. Am.*, *49*, 538-540.

Araseki, T., Ozawa, K., Ono, S., & Ochiai, K. (1983). Multi-pulse excited speech coder based on maximum crosscorrelation search algorithm. *Proceedings of IEEE Global Telecommunications Conference*, 23.3.

Atal, B. S., & Schroeder, M. R. (1979). Predictive coding of speech signals and subjective error criteria. *IEEE Transactions on Acoustics, Speech and Signal Processing*, *ASSP-27*, 247-254.

Atal, B. S., & Remde, J. R. (1982). A new model of LPC excitation for producing natural sounding speech at low bit rates. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, 614-617.

Atal, B. S. (1987). Stochastic gaussian model for low-bit rate coding of LPC area parameters. *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing*, 51.1.

Borden, G. J., & Harris, K. S. (1984). *Speech science primer* (second edition). Baltimore, MD: Williams & Wilkins.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*, 254-259, Hillsdale, NJ: Lawrence Erlbaum Associates.

Egan, J. P. (1948). Articulation testing methods. *Laryngoscope*, *58*, 955-991.

Fairbanks, G. (1958). Test of phonemic differentiation: The rhyme test. *J. Acoust. Soc. Am.*, *30*, 596-600.

Flege, J. E. (1987). The production and pereption of foreign language speech sounds. In H. Winitz (Ed.), *Human communication and its disorders, vol. 1*, Norwood, NJ: Ablex Publishing.

Fujimura, O. (1962). Analysis of nasal consonants. *J. Acoust. Soc. Am.*, *34*, 1865-1875.

Gaies, S. J., Gradman, H. L., & Spolsky, B. (1977). Toward the measurement of functional proficiency: Contextualization of the noise test. *TESOL Quarterly*, *11*, 51-57.

Gat, I. B., & Keith, R. W. (1978). An effect of linguistic experience. *J. Audiology*, *17*, 339-345.

Goodman, D. J., Goodman, J. S., & Chen, M. (1978). Intelligibility and ratings of digitally coded speech. *IEEE Transactions on Acoustics, Speech and Signal Processing*, *ASSP-26*, 5, 403-409.

Greene, B. G., Manous, L. M., & Pisoni, D. B. (1984). Perceptual evaluation of DECtalk: A final report on Version 1.8. In *Research on speech perception progress report no. 10*. Bloomington, IN: Speech Research Laboratory, Psychology Department, Indiana University.

Greene, B. G., Logan, J. S., & Pisoni, D. B. (1986). Perception of synthetic speech produced automatically by rule: Intelligibility of eight text-to-speech systems. *Behavior Research Methods, Instruments, & Computers*, *18*, 100-107.

Greene, B. G. (1986). Perception of synthetic speech by nonnative speakers of English. *Proceedings of the Human Factors Society*, *2*, 1340-1343.

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, *1*, 1-7.

Hawkins, S., & Stevens, K. N. (1985). Acoustic and perceptual correlates of the non-nasal-nasal distinction for vowels. *J. Acoust. Soc. Am.*, *77*, 1560-1575.

House, A. S., & Stevens, K. N. (1956). Analog studies of the nasalization of vowels. *J. Speech and Hearing Disorders*, *21*, 218-232.

House, A. S., Williams, C. E., Hecker, M. H., & Kryter, K. D. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. *J. Acoust. Soc. Am.* *37*, 158-166.

Itakura, F., & Saito, S. (1970). A statistical method for estimation of speech spectral density and formant frequencies. *Elec. and Comm. in Japan*, *53-A*, 36-43.

Kahn, M., & Garst, P. (1983). The effects of five voice characteristics on LPC quality. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, 531-534.

Kalikow, D. N., Huggins, A. W., Blackman, E., Vishu, R., & Sullivan, F. (1976). Speech intelligibility and quality measurement. In *Speech Compression Techniques for Secure Communication*, *BBN Report No. 3226*.

Logan, J. S., Pisoni, D. B., & Greene, B. G. (1985). Measuring the segmental intelligibility of synthetic speech: Results from eight text-to-speech systems. *Research on speech perception progress report no. 11*. Bloomington, IN: Speech Research Laboratory, Psychology Department, Indiana University.

Mack, M. (1987). Peception of natural and vocoded sentences among English monolinguals and German-English bilinguals. *J. Acoust. Soc. Am. Suppl. 1*, *81*, A16.

Makhoul, J. (1975). Spectral linear prediction: properties and applications. *IEEE Transactions on Acoustics, Speech and Signal Processing*, *ASSP-23*, 283-296.

Malecot, A. (1960). Vowel nasality as a distinctive feature in American English. *Language*, *36*, 222-229.

Markel, J. D., & Gray, Jr., A. H. (1976). *Linear prediction of speech*. New York: Springer-Verlag.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *J. Acoust. Soc. Am. 27*, 338-352.

Nakatani, L. H., & Dukes, K. D. (1973). A sensitive test of speech communication quality. *J. Acoust. Soc. Am. 53*, 1083-1092.

Nooteboom, S. G., & Doodeman, G. J. N. (1980). Word recognition from fragments of spoken words by native and non-native listners. *IPO Annual Progress Report*, *15*, 42-47.

Nusbaum, H. C., Dedina, M. J., & Pisoni, D. B. (1984). Perceptual confusions of consonants in natural and synthetic CV syllables. *Speech research lab. tech. note. 84-02*. Bloomington, IN: Speech Research Laboratory, Psychology Department, Indiana university.

Nye, P. W., & Gaitenby, J. H. (1973). Consonant intelligibility in synthetic speech and in a natural speech control (Modified Rhyme Test results). *Status report on speech research SR-33*, (pp. 77-91). New Haven, CT: Haskins Laboratories.

Ozawa, K., & Araseki, T. (1986). High quality multi-pulse speech coder with pitch prediction. *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing*, 1689-1692.

Ozawa, K., Ono, S., & Araseki, T. (1986). A study on pulse search algorithms for multipulse excited speech coder realization. *IEEE Journal on Selected Areas in Communications*, *SAC-4*, 133-141.

Pisoni, D. B. (1978). Speech Perception. In W. K. Estes (Ed.) *Handbook of learning and cognitive processes, vol. 6*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Pisoni, D. B. (1979). Some measures of intelligibility and comprehension. In (J. Allen, S. Hunnicut, & D. H. Klatt, Eds.) *Conversion of unrestricted text to speech*, (Notes for MIT Summer Course 6.69s, July, 1979).

Pisoni, D. B., & Hunicutt, S. (1980). Perceptual evaluation of MITalk: The MIT unrestricted text-to speech system. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, 572-575.

Pisoni, D. B., & Koen, E. (1981). Some comparisons of intelligibility of synthetic and natural speech at different speech-to-noise ratios. In *Research on speech perception progress report no. 7*. Bloomington, IN: Speech Research Laboratory, Psychology Department, Indiana University.

Pisoni, D. B. (1982). Perception of speech: The human listener as a cognitive interface. *Speech Technology*, *1*, April, 10-23.

Pisoni, D. B., Nusbaum, H. C., Luce, P. A., & Schwab, E. C. (1983). Perceptual evaluation of synthetic speech: some considerations of the user/system interface. *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing*, 553-538.

Pisoni, D. B., Nusbaum, H. C., & Greene, B. G. (1985). Perception of synthetic speech generated by rule. *Proceedings of The IEEE*, *73*, 1665-1676.

Pisoni, D. B., Manous, L. M., & Dedina, M. J. (1986). Comprehension of natural and synthetic speech: 2. Effects of predictability on verification of sentences controlled for intelligibility. *Research on speech perception progress report no. 12*, Bloomington, IN: Speech Research Laboratory, Psychology Department, Indiana University.

Pisoni, D. B., & Dedina M. J. (1986). Comprehension of digitally encoded natural speech using a sentence verification task (SVT): A first report. *Research on speech perception progress report no. 12*. Bloomington, IN: Speech Research Laboratory, Psychology Department, Indiana University.

Pisoni, D. B., & Luce, P. A. (1986). Speech perception: research, theory, and the principal issues. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines, vol 1*, Orlando, FL: Academic Press.

Preusse, J. W. (1969). The consonant reognition test. *Command Res. Develop. Tech. Rep. ECOM 3207*.

Rabiner, L. R., & Schafer, R. W. (1978). *Digital processing of speech signals*, Englewood Cliffs, NJ: Prentice-Hall.

Schmidt-Nielsen, A., & Kallman, H. J. (1987). Response time to a sentence verification task as a function of LPC narrow-band processing and bit error rate. *J. Acoust. Soc. Am.*, *Suppl. 1*, 81, A15.

Schroeder, M. R., & Atal, B. S. (1985). Code-excited linear prediction (CELP): High quality speech at very low bit rates. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, 25.1.

Singhal, S., & Atal, B. S. (1984). Improving performance of multi-pulse LPC coders at low bit rates. *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing*, 1.3.

Spolsky, B., Sigurd, B, Satao, M., Walker, E., & Arterburn, C. (1968). Preliminary studies in the development of techniques for testing overall second language proficiency. *Language Learning*, *18* (Special Issue No. 3), 79-101.

Stevens, K. N., Libermann, A. M., Studdert-Kennedy, M., & Ohman, S. E. G. (1969). Crosslanguage study of vowel perception. *J. Language and Speech*, *12*, 1-23.

Transco, I. M., & Atal, B. S. (1986). Efficient procedures for finding the optimum innovation in stochastic coding. *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing*, 44.5.

Voiers, W. D., Cohen, M. F., & Mickunas, J. (1965). Evaluation of speech processing devices, I. Intelligibility, quality, speaker recognizability. *Final Report, Contract No. AF19(628)4195*, OAS.

Wang, M. D., & Bilger, R. C. (1973). Consonant confusions in noise: a study of perceptual features. J. Acoust. Soc. Am. 54, 1248-1266.

Wong, D. Y., & Markel, J. D. (1978). An intelligibility evaluation of several linear prediction vocoder modifications. IEEE Transactions Acoustics, Speech and Signal Processing, ASSP-26, 424-435.

Yuchtman, M., Nusbaum, H. C., & Pisoni, D. B. (1985). Consonant confusions and perceptual spaces for natural and synthetic speech. J. Acoust. Soc. Am., Suppl. 1, 78, NN12.

Reference Note

Generally, in an analysis of variance, we assume that changes in stimuli cause uniform changes in behaviors of subjects. That assumption is reasonable when subjects are drawn from one uniformly distributed population, such as native listeners. However, the assumption may not be reasonable when two different groups of listeners, such as native and non-native listeners are used. In such a case, a nonlinear transformation of the data may be appropriate. Even more important, the unit of measurement for the proportions may not be constant over the measurement scale, especially at the endpoints of the scale. In the case of the present experiment, the error data obtained in the various conditions differed only a small amount at the initial portion of the measurement scale. Cohen and Cohen (1975) argue that differences at the endpoints of a measurement scale, such as percent error, are more important than differences in the middle of the scale. Thus, the difference between conditions with 2% and 4% error is more important than the difference between 52% and 54% error since 4% is twice as large as 2% whereas 54% is only fractionally larger than 52%. Therefore, it seemed appropriate to use a nonlinear transformation on our data to emphasize differences that occured at the endpoints of the measurement scale. We carried out an analysis of variance on the error data obtained from native and non-native listeners using the nonlinear arcsine transformation (Cohen & Cohen, 1975). The arcsine transformation is defined as

$$A = 2 \text{ arcsine} \sqrt{p}\,, \qquad (2)$$

where p is a proportion and A is a transformed value (measured in radians). The results of the analysis of variance using the transformed data were the same as the analysis using the untransformed data.

[RESEARCH ON SPEECH PERCEPTION Progress Report No. 13 (1987) Indiana University]

F1 Structure Provides Information for Final-Consonant Voicing*

W. Van Summers

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, IN 47405

---

*This research was supported by NIH Training Grant NS-07134-09 to Indiana University in Bloomington. Thanks to David Pisoni, John Mullennix, and Chris Martin for comments on an earlier version of this paper.

Abstract

Previous research has shown that F1 offset frequencies are generally lower for vowels preceding voiced consonants than for vowels preceding voiceless consonants. Furthermore, it has been shown that listeners use these differences in offset frequency in making judgments about final-consonant voicing. A recent production study (Summers, 1987) reported that F1 frequency differences due to postvocalic voicing are not limited to the final transition or offset region of the preceding vowel. Vowels preceding voiced consonants showed lower F1 onset frequencies and lower F1 steady-state frequencies than vowels preceding voiceless consonants. The present study examined whether F1 frequency differences in the initial transition and steady-state regions of preceding vowels effect final-consonant voicing judgments in perception. The results suggest that F1 frequency differences in these early portions of preceding vowels do, in fact, influence listeners' judgements of postvocalic consonantal voicing.

F1 Structure Provides Information for Final-Consonant Voicing

It is well-known that the voicing feature of a postvocalic consonant has predictable effects on the temporal and spectral structure of a preceding vowel. Specifically, vowels preceding voiced consonants will generally have longer durations than vowels preceding voiceless consonants (House and Fairbanks, 1953; House, 1961; Luce and Charles-Luce, 1985; Mack, 1982). In addition, final-consonant voicing has an influence on first formant final transition (F1FT) characteristics of preceding vowels. Vowels preceding voiced consonants generally contain falling F1FT's, with F1 offset frequencies well below F1 steady-state frequencies. Vowels preceding voiceless consonants may not contain F1FT's, with F1 maintaining its steady-state frequency until vowel offset (Walsh and Parker, 1983). When these vowels do contain F1FT's, these final transitions are generally brief, terminating at higher offset frequencies than F1FT's for vowels preceding voiced consonants (Hillenbrand, Ingrisano, Smith, and Flege, 1984; Summers, 1987; Wolf, 1978).

Perceptual research has shown that preceding vowel duration supplies useful information to listeners concerning final-consonant voicing. Long vowel durations cue voiced final consonants and short vowel durations cue voiceless final consonants (Denes, 1955; Raphael, 1972). F1FT characteristics have also been shown to influence final-consonant voicing decisions. When vowel durations are approximately equal, utterances containing falling F1FT's and low F1 offset frequencies are judged as ending in voiced consonants more often than utterances without F1FT's or with gradual F1FT's which terminate at higher frequencies (Hillenbrand et al., 1984; Walsh and Parker, 1983; Wolf, 1978).

A recent study examining the effects of final-consonant voicing on vowel production (Summers, 1987) showed that voicing-related differences in F1 frequency are not limited to F1 final transition regions. As in previous studies, Summers found that vowels preceding voiceless final consonants had higher F1 offset frequencies than vowels preceding voiced consonants. However, final-consonant voicing also infuenced F1 frequencies during initial transition and steady-state portions of the preceding vowel. These F1 frequency differences were consistent and reliable for each of three speakers. Similar data regarding voicing effects on F1 steady-state frequency have been reported previously (Wolf, 1978; Revoile, Pickett, Holden, and Talkin, 1982).

The present study examined whether differences in F1 frequency in the initial-transition and steady-state portions of preceding vowels provide perceptual information about postvocalic voicing. Results of previous perceptual studies support the hypothesis that final-consonant voicing information is present in early portions of preceding vowels. Using truncated stimuli, equated for duration and containing no final formant transitions, Wolf (1978) reported evidence of final-consonant voicing information present in the initial 50 ms of preceding vowels. Similar findings were reported by O'Kane (1978). There is some evidence that F1 frequency differences prior to F1 final transition onset may provide some of this early final-consonant voicing information. Mermelstein (1978) collected final-consonant voicing judgments for stimuli which varied in steady-state vowel duration and steady-state F1 frequency. Consistent with Summers' (1987) production data, Mermelstein reported that high F1 steady-state frequencies were associated with an increase in voiceless final-consonant judgments. A major focus of the present study was to explicitly examine F1 steady-state frequency as a potential voicing cue.

The stimuli used in the present experiment also allowed an examination of several other potential sources of final-consonant voicing information. The stimuli varied in F1 onset frequency, F1 steady-state frequency, F1FT slope, F1 offset frequency, and total vowel duration. As mentioned earlier, previous research has clearly demonstrated that vowel duration supplies important final-consonant voicing information. In addition, F1FT slope and F1 offset frequency have previously been proposed as sources of final-consonant voicing information (Walsh and Parker, 1983; Walsh, Parker, and Miller, 1987; Wolf, 1978). The present study allowed an examination of each of these potential sources of voicing information and provided a test of whether F1 onset frequency and F1 steady-state frequency also contribute final-consonant voicing information in perception.

Method

Subjects. Thirty-two Indiana University undergraduate students participated as subjects to fulfill course requirements in Introductory Psychology. All subjects were native speakers of American English with no reported history of a speech or hearing disorder at the time of testing. Subjects were randomly assigned to one of two response conditions, to be described below. Seventeen subjects participated in the two-alternative response condition and 15 subjects participated in the four-alternative response condition.

Stimuli. Six series of b-vowel-consonant syllables were synthesized using the cascade formant synthesis software developed by Klatt (1980). All stimuli contained an initial 10 ms burst and initial formant transitions appropriate to the labial stop consonant /b/. Three series contained formant values appropriate to the vowel /a/ (the /a/ series) and three series had formant values appropriate to /æ/ (the /æ/ series). Within each series, six stimuli were created by increasing vowel duration in 35 ms steps from 115 ms to 290 ms. Vowel duration was manipulated through iteration of steady-state regions. All stimuli contained final formant transitions appropriate to the stop consonants /b/ and /p/. All stimuli were composed of an initial burst, an initial formant transition region, a steady-state region, and a final formant transition region. As already mentioned, stimuli within a series varied in total vowel duration from 115 ms to 290 ms. Total vowel duration includes initial transition, steady-state, and final transition regions. Stimuli within a series will henceforth be referred to in terms of their total vowel duration. Thus, the briefest member of each series will be referred to as the 115 ms member. The total duration of each stimulus is actually 10 ms greater than the total vowel duration due to the initial burst.

Synthesis parameters for the 115 ms member of each series are listed in the Appendix. With the exception of F1, all stimuli based on a given vowel used identical parameters. Figure 1 shows the F1 trajectory for the 115 ms member of each series. Stimuli from the three /a/-vowel series are shown in the upper panel of the figure and stimuli from the /æ/-vowel series are shown in the lower panel.

Insert Figure 1 about here

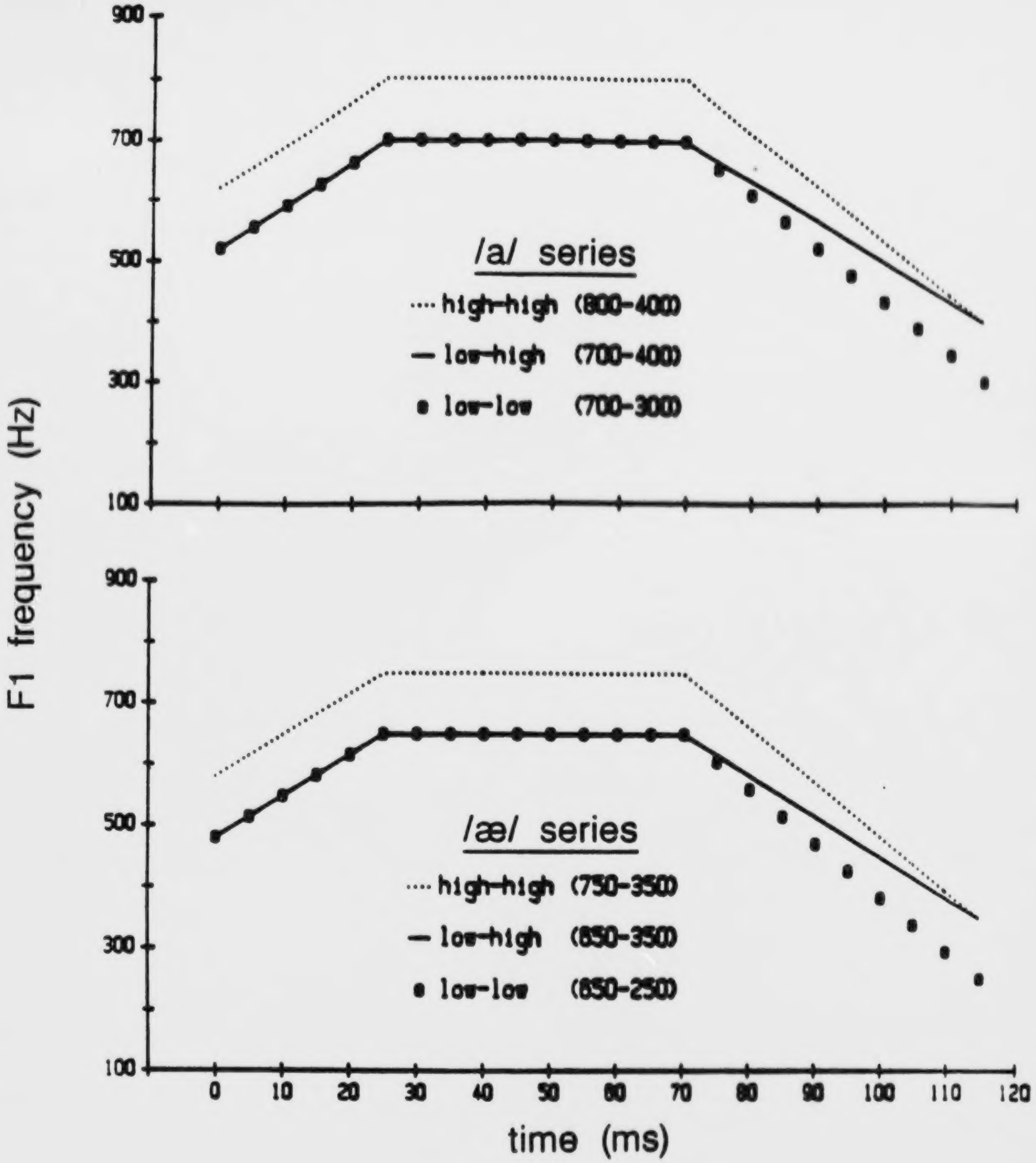


Figure 1. F1 frequency trajectory for the 115 ms member of each stimulus series. Upper panel shows stimuli from /a/-vowel series. Lower panel shows stimuli from /æ/-vowel series.

The three series of stimuli based on a given vowel differed in terms of F1 onset frequency, F1 steady-state frequency, F1FT slope, and F1 offset frequency. F1 onset frequency and F1 steady-state frequency covaried so that stimuli which differed in onset frequency differed in steady-state frequency by the same amount. Since F1 onset frequency and F1 steady-state frequency were correlated in this manner, they will generally be referred to as one variable: ON+SS frequency. The series are labeled in terms of F1 steady-state frequency and F1 offset frequency. For example, in the upper panel of Figure 1, the stimulus from the 800-400 series has an 800 Hz steady-state frequency and a 400 Hz offset frequency. The three /a/ series were: 800-400, 700-400, and 700-300. The /æ/ series were: 750-350, 650-350, and 650-250. For each vowel, there were three types of series: high-high series, which contain a high F1 ON+SS frequency and a high F1 offset frequency; low-high series which contain a low F1 ON+SS frequency and a high F1 offset frequency; and low-low series which contain a low F1 ON+SS frequency and a low F1 offset.

Procedure. Stimuli were presented at 70 dB SPL over matched and calibrated TDH-39 headphones. Stimulus presentation was controlled by a PDP 11/34 computer. Identification responses for the six series of stimuli were collected in two different testing conditions. In each condition, 4 blocks of 144 trials were presented with a 3 sec inter-trial interval. In the four-alternative condition, stimuli from all series were randomized as a group and presented to subjects. In this condition, subjects identified the stimuli as "bob," "bop," "bab," or "bap" in a four-alternative forced-choice ID task. Each of the 36 stimuli were presented 4 times in each block for a total of 16 responses per stimulus.

In the two-alternative condition, the /a/ and /æ/ series were randomized separately and presented in alternating blocks. Subjects labeled stimuli as "bob" or "bop" in blocks containing stimuli from the /a/ series (blocks 1 and 3) and as "bab" or "bap" in blocks containing stimuli from the /æ/ series (blocks 2 and 4). Each stimulus was presented 8 times per block in 2 blocks for a total of 16 responses per stimulus. In each testing condition, identification responses were made by pressing the appropriate button on a response box placed directly in front of the subject. Response boxes containing four buttons were used in the four-alternative condition; boxes containing two buttons were used in the two-alternative condition.

Two testing conditions (two- and four-alternative) were included in the present experiment for several reasons. The two-alternative task has generally been used in previous studies when a binary decision (e.g., voiced/voiceless, stop/continuant) is required. It is the simplest task available for testing whether differences in F1 characteristics in the present stimuli influenced final-consonant voicing judgments. The four-alternative task requires a vowel response in addition to a final-consonant response. Therefore, it is a more complex task with greater stimulus uncertainty than the two-alternative task in which the vowel is constant within a block of trials. A comparison of performance in the two testing conditions provided an indication of whether any effects of F1 structure on voicing judgments were consistent across testing conditions or if these effects were conditioned by the predictability of the surrounding context. The four-alternative condition also allowed a verification that the stimuli were unambiguous in terms of vowel (/a/ or /æ/).

Results

The results from the four-alternative condition were examined to assure that subjects were correctly identifying stimulus vowels. Four of the 15 subjects in this condition identified the vowel correctly 100% of the time. The poorest performance by any subject involved 14 vowel errors out of 576 responses (97.6% correct vowel identification). Across subjects, mean percentage of correct vowel responses was 99.4%. Trials in which vowel errors occurred were excluded from further analysis.

For each subject, the percentage of /bab/ or /bæb/ responses to each member of a given series was calculated. The best-fitting normal ogive through these points was then determined (Woodworth, 1938). The 50% point of this ogive was taken as the crossover point in the labeling function: the vowel duration at which final /b/ and final /p/ responses were equally likely. These 50% crossover points were used as dependent measures in an analysis of variance with vowel (/a/ versus /æ/) and series type (high-high, low-high, or low-low) as within-subjects factors and with response condition (four-alternative versus two-alternative) as a between-subjects factor.

Response condition (four-alternative versus two-alternative) did not have a significant influence on crossover durations. Mean crossover durations were 199.0 ms in the two-alternative condition and 193.9 ms in the four-alternative condition ($F(1,30) = 1.23$, $p = .277$). No significant interactions involving response condition were obtained.

Vowel identity (/a/ versus /æ/) and series type (high-high, low-high, or low-low) both had significant effects on crossover durations. Mean crossover durations are broken down by vowel and series type in Table 1. Given the lack of any significant effect of response condition on crossover durations, the values reported in Table 1 are collapsed across response conditions.

Insert Table 1 about here

Vowel identity had a significant main effect on crossover duration. Series based on /a/ displayed longer crossover durations than series based on /æ/ ($F(1,30) = 4.82$. $p = .036$). This pattern was consistent for each of the three series types (see Table 1). None of the interactions involving vowel condition approached significance.

Finally, the analysis of variance demonstrated a clear effect of series type (high-high, low-high, or low-low) on crossover vowel duration ($F(2,60) = 24.58$, $p < .0001$). Mean crossover durations were greater in high-high series than in low-high series and greater in low-high series than in low-low series. This pattern was consistent for both /a/ and /æ/ series (see Table 1). None of the interactions involving series type approached significance. The significant main effect of series type suggests that one or more of the differences in F1 structure between the three types of series influenced subjects' judgments of final-consonant voicing. The F1 characteristics responsible for this significant effect were then examined in a more fine-grained analysis.

Table 1

Mean crossover durations in ms collapsed across response conditions

	Series Type			
Vowel	high-high	low-high	low-low	Mean
/a/	213.6	198.8	188.8	200.4
/æ/	204.6	188.9	184.8	192.8
Mean	209.1	193.9	186.8	

The three series of stimuli created for each vowel afford three pairwise comparisons of identification performance. Each of these pairwise comparisons involve series contrasting in different F1 characteristics. Each of these comparisons will now be described. Because no significant effect of response condition was obtained in the analysis of variance, the data were collapsed across response conditions in making these comparisons between series types.

Figure 2 displays the identification data for each of the six stimulus series. This figure contains the relevant data for each of the pairwise comparisons described below. The results for the /a/ series are plotted in the upper panel of the figure and the results for the /æ/ series appear in the lower panel. Mean percentage of /bab/ (upper panel) and /bæb/ (lower panel) responses to each stimulus are shown and the best-fitting normal ogives through these means are plotted (Woodworth, 1938).

Insert Figure 2 about here

High-high versus Low-low series

The first comparison examined was between the high-high series and the low-low series for each vowel. This comparison involved series with equal F1 final transition slopes but which contrasted in F1 ON+SS frequency and F1 offset frequency. Examining the data for high-high and low-low series in Figure 2, it can be seen that stimuli from low-low series received more final /b/ responses than stimuli from high-high series. That is, stimuli with low F1 ON+SS frequencies and low F1 offset frequencies received more final /b/ responses. This pattern was consistent at every vowel duration for both the /a/ and /æ/ vowel series. Planned-comparisons of mean crossover points for high-high series versus low-low series demonstrated that these differences were significant (Dunn's multiple comparison procedure (Kirk, 1982)). Stimuli from low-low series were identified as ending in /b/ at shorter vowel durations than stimuli from high-high series ($tD(60) = 4.88$, $p < .01$).

The high-high versus low-low series data suggest that judgments of final-consonant voicing were influenced by F1 onset frequencies, F1 steady-state frequencies, F1 offset frequencies, or a combination of these cues. These results are consistent with the hypothesis that high F1 ON+SS frequencies cue voiceless final consonants for these vowels. However, the results do not provide unequivocal support for F1 onset frequencies or F1 steady-state frequencies as voicing cues, because F1 offset frequency differences alone may explain the observed pattern.

High-high versus Low-high series

The second pairwise comparison available in these data involves stimulus series in which F1 ON+SS frequency differences are not confounded with offset frequency differences. This second comparison is between high-high and low-high series. These series had equal F1 offset frequencies but differed in F1 ON+SS frequency and in F1 final transition slope.

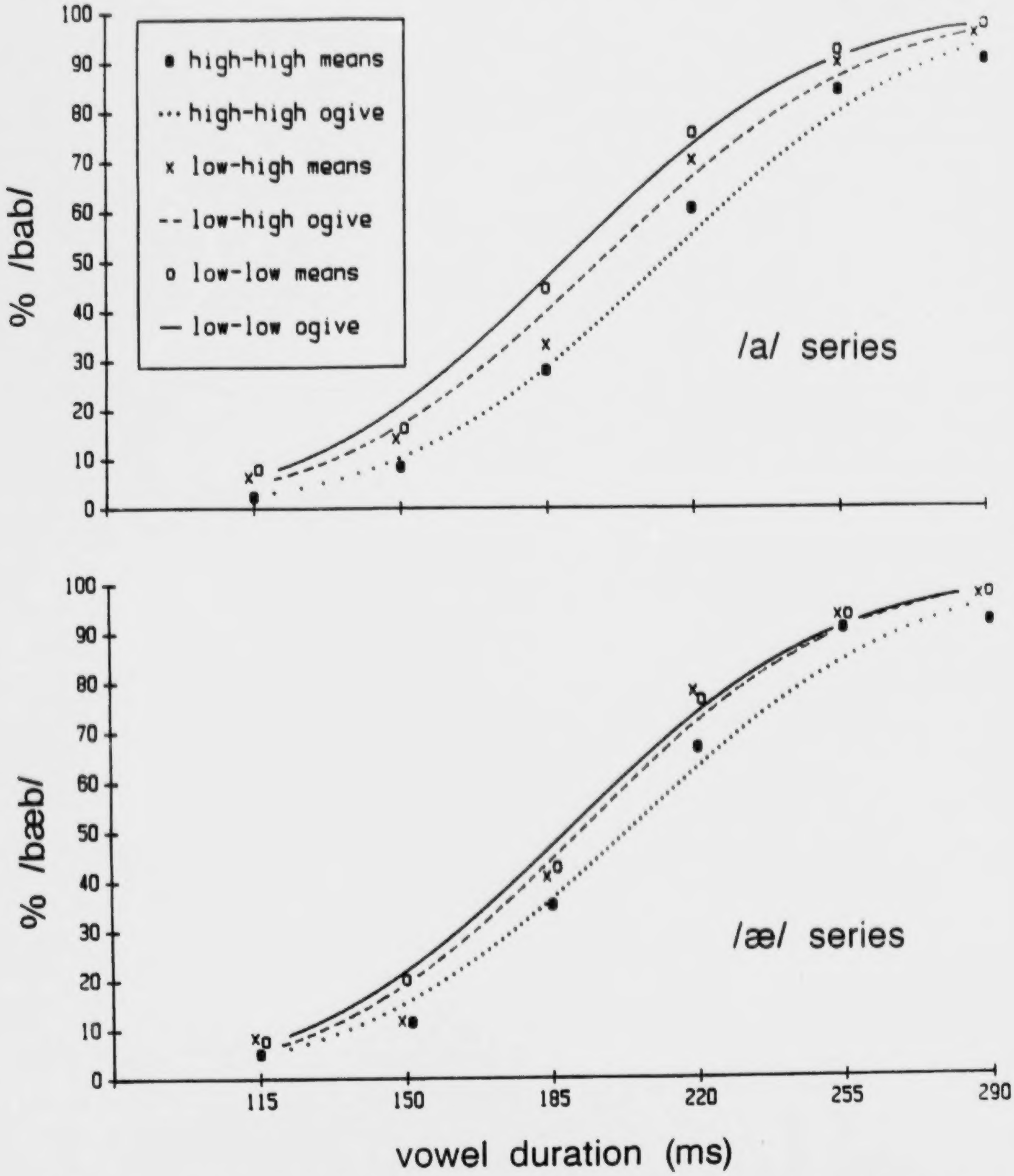


Figure 2. Mean percentage of voiced final-consonant judgments at each vowel duration for each vowel series. Best-fitting normal ogives through the mean values for each series are also shown. Upper panel shows /a/-vowel series and lower panel shows /æ/-vowel series. Mean values are slightly displaced horizontally when necessary to maintain clarity.

As in the previous comparison, stimuli containing low F1 ON+SS frequencies were more likely to be identified as ending in a voiced final consonant than stimuli with high F1 ON+SS frequencies. This pattern was consistent at every vowel duration for both vowels (see Figure 2). Planned comparisons of mean crossover durations for high-high versus low-high series demonstrated that these differences were significant. Stimuli from low-high series were identified as ending in /b/ at shorter vowel durations than stimuli from high-high series (tD(60) = 3.34, p < .01). These results demonstrate a significant effect of F1 ON+SS frequency on voicing judgments in the absence of F1 offset frequency differences.

Walsh et al. (1987) have suggested that F1 final transition slopes may provide final-consonant voicing information with steeper slopes cuing voiced final consonants. The present data for high-high and low-high series do not appear to support this hypothesis. In these data, stimuli from high-high series contain steeper final transitions than stimuli from low-high series, but are more often judged to contain a voiceless final consonant. It may be that the effect of final transition slope is masked, in this case, by the greater effect of F1 ON+SS differences. Alternatively, the discrepancy between the present data and the Walsh results may be due to the confounding of F1 final transition slope and F1 offset frequency in the Walsh et al. study. This point will be returned to in the discussion below. A comparison between stimulus series very similar to those used by Walsh et al. (1987) is available in the present study and is described next.

Low-high versus Low-low series

The final pairwise comparison involves stimuli from low-high versus low-low series. For a given vowel, stimuli from these two series contained equal F1 onset frequencies and F1 steady-state frequencies but differed in terms of F1 final transition slope and F1 offset frequency. Low-low series stimuli contained steeper F1 final transitions and lower F1 offset frequencies than low-high series stimuli (see Figure 1).

Examining performance for low-high and low-low series in Figure 2, it can be seen that stimuli containing lower F1 offset frequencies and steeper final transitions (i.e., low-low series stimuli) received more final /b/ responses than stimuli with higher F1 offsets and more gradual F1 final transitions (i.e., low-high series stimuli). This overall pattern can be seen by comparing the ogives in each panel of the figure. However, the effect is not as consistent as in the earlier comparisons, particularly for stimuli based on /æ/. In the lower panel of Figure 2, which displays data for series based on /æ/, there are mean values that are not consistent with the overall pattern of results. For example, at vowel duration = 220 ms, the low-high stimulus received more voiced responses than its 220 ms low-low counterpart. These reversals on the overall pattern did not occur in the earlier comparisons. Planned comparisons of mean crossover points for low-high versus low-low series were not statistically significant (tD(60) = 1.55, N.S.). Thus, it appears that the earlier comparisons, both of which involved series differing in F1 ON+SS frequency, demonstrated more consistent effects on voicing judgments than the final comparison in which F1 onset frequencies and F1 steady-state frequencies did not vary.

The low-high versus low-low stimulus comparison provides a fairly close replication of an earlier study by Walsh et al. (1987). In that study, stimuli with steeper F1 final transition slopes and lower F1 offset frequencies received significantly more voiced final-consonant judgments than stimuli with more gradual F1 final transition slopes and higher F1 offset frequencies. The present results only partially replicate these earlier findings. While the general pattern of results for the low-high versus low-low series is consistent with the pattern reported by Walsh et al. (1987), low-high versus low-low crossover points did not significantly differ in the present study. The Walsh et al. (1987) study and its conclusions will be taken up in the discussion below.

Discussion

The ogives plotted in the panels of Figure 2 show a consistent pattern of results for both /a/ and /æ/. The largest change in voicing decisions involved high-high versus low-low series. These were series in which both ON+SS frequency and offset frequency differences were present. A smaller change was seen in the high-high versus low-high comparison. These series contained the ON+SS frequency differences present in the previous comparison but did not differ in offset frequency. The significant difference in crossover durations for high-high versus low-high series suggests an effect of F1 ON+SS frequency on voicing decisions independent of F1 offset frequency. The fact that a larger change in voicing decisions was present in the high-high versus low-low comparison than in the high-high versus low-high comparison suggests an effect of F1 offset frequency on voicing judgments which is independent of the ON+SS frequency effect. In short, the results suggest that both F1 ON+SS frequency and F1 offset frequency provide perceptual information for final-consonant voicing. For the vowels examined, low F1 ON+SS frequencies and low offset frequencies tended to produce voiced final-consonant judgments.

It could be argued that the present results do not support F1 offset frequency as a voicing cue since the results of the low-high versus low-low comparison, in which offset frequency differences were present, were not statistically significant. As a result, the data do not provide strong support for F1 offset frequency as voicing cue. However, the present results are not inconsistent with previous work in which F1 offset frequencies have appeared to provide voicing information (Wolf, 1978; Hillenbrand et al., 1984). The results of the present low-high versus low-low comparison, while not statistically significant, were in the expected direction based on this previous work. Stimuli with low F1 offset frequencies tended to receive more voiced final-consonant judgments than stimuli with high F1 offsets. The pattern was consistent at every vowel duration for /a/ (see Figure 2). While the data were less consistent for /æ/, the overall pattern was again in the expected direction. The variability in the data for /æ/ appears to be the cause of the overall lack of statistical significance.

Additional evidence that high F1 onset, steady-state, and offset frequencies cue voiceless final consonants comes from comparing labelling performance for /a/ versus /æ/. First it should be noted that these vowels differ in inherent duration. According to Peterson and Lehiste's (1960) measurements, /æ/ has the longest inherent duration of all English monophthongs, considerably longer than /a/. If listeners adjust their perceptual judgments for these inherent durational differences, it would be expected that judgments of final-consonant voicing would switch from voiceless to voiced at briefer durations for the /a/ series than the /æ/ series.

However, in the present data, mean crossover points were significantly earlier for /æ/ than /a/. This unexpected result may be due to frequency differences between the /a/ and /æ/ series as synthesized. F1 frequencies at onset, steady-state, and offset were higher for /a/ stimuli versus /æ/ stimuli when matching series are compared (e.g., high-high versus high-high). The higher F1 frequencies used in synthesizing the /a/ series may have encouraged listeners to hear these stimuli as ending in voiceless consonants more often than stimuli from /æ/ series which contained lower F1 frequencies.

There is little evidence that F1 final transition slope had a consistent influence on voicing judgments in this study. Consider the data for the high-high and low-high series. In this comparison, high-high stimuli contain steeper F1 final transitions than low-high stimuli. If, as Walsh et al. (1987) suggest, steep F1FT's cue voiced final consonants, high-high stimuli should be judged as ending in a voiced consonant more often than low-high stimuli. However, exactly the opposite result was observed; high-high stimuli received more voiceless responses than low-high stimuli. If steep F1 final transitions were cuing voiced final consonants in these stimuli, the effect was clearly much weaker than the effect of F1 ON+SS differences and, as a result, was completely masked. These results are consistent with Summers' (1987) production data in which high F1 onset frequencies and high F1 steady-state frequencies are associated with the production of voiceless final consonants. Furthermore, Summers' (1987) data failed to show significant differences in F1FT slope for utterances contrasting in final-consonant voicing.

The data from the low-high versus low-low comparison are much more consistent with the Walsh et al. (1987) hypothesis concerning F1 transition slopes than the data from the high-high versus low-high comparison. However, the low-high versus low-low results provide little support for the Walsh et al. position for two reasons. First, the change in responses for low-high versus low-low series was not statistically significant. Second, even if significant, the results are ambiguous as to the cuing value of final transition slope because, as in the stimuli used by Walsh et al., F1 final transition slope and F1 offset frequency were confounded in this comparison. Lower F1 offset frequencies rather than steeper F1 final transitions may cue voiced consonants in this case.

Since F1 onset frequency and F1 steady-state frequency were correlated in this study, the results do not directly address the relative contribution of onset frequency and steady-state frequency to final voicing decisions. However, evidence from other studies suggests that steady-state frequency may outweigh onset frequency in conveying final voicing information. First, Summers' (1987) production data showed larger differences in F1 steady-state frequency than in F1 onset frequency for utterances contrasting in final consonant voicing. If final-consonant voicing generally influences F1 steady-state frequencies more than F1 onsets, steady-state frequency differences may be more salient and may be relied on more by listeners in making voicing decisions. Second, perceptual experiments examining the effects of selectively deleting portions of vowels on final voicing decisions (Wardrip-Fruin, 1982) have shown that deleting later-occurring portions of vowels has a greater effect on judgments than deleting earlier portions. This finding clearly suggests an important role for final formant transitions in cuing final-consonant voicing. However, it also consistent with the hypothesis that later-occurring steady-state formant regions carry more final-consonant voicing information than initial formant transitions. Finally, a third piece of evidence that steady-state frequency may outweigh onset frequency as a voicing cue in the present study has to do with the durations of initial transitions and steady-state regions in the experimental stimuli. F1 initial transitions were 25 ms long for all stimuli in the present study while steady-state durations varied from 45 ms to 220 ms within each series. The longer durations of steady-state regions relative to initial transitions may have made steady-state frequency differences more salient than frequency differences during initial transitions. Summers' (1987) earlier production data verifies that F1 steady-state regions are generally longer than F1 initial transitions for consonant-vowel-consonant utterances.

The results of the present study contrast in an interesting way with previous work examining the influence of linguistic stress on formant frequencies and vowel durations. Vowel durations are generally longer in stressed utterances than in unstressed utterances (Cooper, Eady, and Mueller, 1985; Parmenter and Trevino, 1936; Summers, 1987). Thus, the presence of stress and the presence of a voiced final consonant both tend to increase vowel duration. However, stress and final-consonant voicing appear to have contrasting influences on F1 frequency. According to the present findings, lower F1 frequencies are more likely to be associated with voiced final consonant judgments than with voiceless consonants judgments. As a result, vowel lengthening due to final-consonant voicing is associated with a lowering of F1. Stress-related vowel lengthening has exactly the opposite effect on F1. For low vowels such as /a/ and /æ/, F1 frequencies are higher in stressed utterances than in unstressed utterances (DeLattre, 1969; Gay, 1978). This suggests that stress-related vowel lengthening may be disambiguated from voicing-related vowel lengthening based on F1 frequency information (see Summers, 1987).

Finally, it should be pointed out that the F1 frequency cues to final-consonant voicing described above may not be equally available for all vowels. The present study examined the low vowels /a/ and /æ/ which contain relatively high F1 frequencies. Previous acoustic measurements showing clear voicing-related differences in F1 frequencies have also tended to focus on vowels containing high first formants (Revoile et al. 1982; Summers, 1987; Wolf, 1978). There is some question as to whether vowels containing lower F1 frequencies would show consistent voicing-related differences in F1 frequencies and whether F1 frequency differences would supply reliable voicing information for these vowels. Acoustic measurements by Hillenbrand et al. (1984) show much larger voicing-related differences in F1 offset frequency for utterances containing /a/ and /æ/ than for utterances containing /i/ and /u/. That is, utterances containing vowels with relatively high F1 frequencies showed larger voicing-related changes in F1 offset frequencies than utterances containing vowel with low F1 frequencies. If Hillenbrand's findings on F1 offset frequencies also hold for F1 onset and steady-state frequencies, it may be that F1 frequency differences play a larger role in cuing final consonant voicing for utterances containing low vowels such as /a/ and /æ/ than for utterances containing high vowels such as /i/ and /u/.

It is possible that larger voicing-related changes in F1 are present for low vowels than high vowels as a result of constraints on the variability of tongue height in the production of high vowels such as /i/ and /u/. These vowels are produced with the tongue high in the oral cavity. For these vowels, further increases in tongue height may not be possible without switching from vowel to fricative production. Decreases in tongue height may also be limited since this would presumably move formant frequencies towards those of more central vowels. The limitations on tongue height may not be as strict for low vowels such as /a/ and /æ/. These vowels are produced with the tongue low in the front cavity which results in high F1 frequencies. Presumably a certain amount of lowering is necessary to disamabiguate /a/ and

/æ/ from more central vowels. However, still more lowering may be possible and this extra lowering may occur when voiceless final consonants follow low vowels. This increased lowering of the tongue may be accomplished by an increase in jaw lowering for these utterances. This increase in jaw lowering for utterances containing voiceless final consonants was reported in Summers' (1987) production study.

Conclusion

The results of the present study suggest that F1 frequency information from the initial transition and steady-state regions of preceding vowels influences judgments of voicing for postvocalic consonants. Low F1 frequencies at vowel onset and during steady-state regions were associated with increases in voiced final-consonant judgments. The results also tended to support previous research which has suggested that low F1 offset frequencies also cue voiced final consonants. Further, the results suggest that listeners may use F1 frequency information to distinguish vowel lengthening due to stress from lengthening due to final-consonant voicing. These findings are based on stimuli containing vowels with high F1 frequencies. It is unclear at present whether F1 frequency differences related to final-consonant voicing are as great or as perceptually informative for vowels containing lower F1 frequencies.

References

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America*, *77*, 2142-2156.

Delattre, P. (1969). An acoustic and articulatory study of vowel reduction in four languages. *International Review of Applied Linguistics*, *7*, 295-325.

Denes, P. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America*, *27*, 761-764.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America*, *63*, 223-230.

Hillenbrand, J., Ingrisano, D. R., Smith, B. L., & Flege, J. E. (1984). Perception of the voiced-voiceless contrast in syllable-final stops. *Journal of the Acoustical Society of America*, *76*, 18-27.

House, A. S. (1961). On vowel duration in English. *Journal of the Acoustical Society of America*, *33*, 1174-1178.

House, A. S., & Fairbanks, G. (1953). The influence of consonantal environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, *25*, 105-113.

Kirk, R. E. (1982). *Experimental design* (Second Edition). California: Wadsworth.

Klatt, D. K. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, *67*, 971-995.

Luce, P. A., & Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *Journal of the Acoustical Society of America*, *78*, 1949-1957.

Mack, M. (1982). Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *Journal of the Acoustical Society of America*, *71*, 173-178.

Mermelstein, P. (1978). On the relationship between vowel and consonant identification when cued by the same acoustic information. *Perception and Psychophysics*, *23*, 331-336.

O'Kane, D. (1978). Manner of vowel termination as a perceptual cue to the voicing status of postvocalic stop consonants. *Journal of Phonetics*, *6*, 311-318.

Parmenter, C. E., & Trevino, S. N. (1936). Relative durations of stressed and unstressed vowels. *American Speech*, *10*, 129-133.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, *32*, 693-703.

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in American English. Journal of the Acoustical Society of America, 51, 1296-1303.

Revoile, S., Pickett, J. M., Holden, L. D., & Talkin, D. (1982). Acoustic cues to final stop consonant voicing for impaired- and normal-hearing listeners. Journal of the Acoustical Society of America, 72, 1145-1154.

Summers, W. V. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. Journal of the Acoustical Society of America, 82, 847-863.

Walsh, T., & Parker, F. (1983). Vowel length and vowel transition: cues to [+/- voice] in postvocalic stops. Journal of Phonetics, 11, 407-412.

Walsh, T., Parker, F., & Miller, C. J. (1987). The contribution of rate of F1 decline to the perception of [+/- voice]. Journal of Phonetics, 15, 101-103.

Wardrip-Fruin, C. (1982). On the status of phonetic categories: Preceding vowel duration as a cue to voicing in final stop consonants. Journal of the Acoustical Society of America, 71, 187-195.

Wolf, C. G. (1978). Voicing cues in English final stops. Journal of Phonetics, 6, 299-309.

Woodworth, R. S. (1938). Experimental Psychology. New York: Holt.

Appendix

Parameter values used in synthesizing the 115 ms member of each series. Parameters held constant for all stimuli were: F0 (120 Hz), F4 (3300 Hz), F5 (3850 Hz), B4 (250 Hz), and B5 (200 Hz).

/a/ series

FRAME	AV	AF	High-high F1	Low-high F1	Low-low F1	F2	F3	AB	B1	B2	B3
0	0	0	200	200	200	1100	2080	0	70	70	120
5	0	0	280	280	280	1113	2173	63	70	70	123
10	0	62	360	360	360	1126	2267	63	70	70	126
15	0	0	440	440	440	1139	2360	63	70	70	129
20	60	0	620	520	520	1152	2453	63	80	70	131
25	60	0	656	556	556	1161	2468	63	90	70	134
30	60	0	692	592	592	1169	2482	63	100	70	137
35	60	0	728	628	628	1178	2497	63	110	70	140
40	60	0	764	664	664	1186	2512	63	120	70	143
45	60	0	800	700	700	1195	2526	0	130	70	146
50	60	0	800	700	700	1203	2541	0	130	70	149
55	60	0	800	700	700	1212	2556	0	130	70	151
60	60	0	800	700	700	1220	2571	0	130	70	154
65	60	0	800	700	700	1220	2585	0	130	70	157
70	60	0	800	700	700	1220	2600	0	130	70	160
75	60	0	800	700	700	1220	2600	0	130	70	160
80	60	0	800	700	700	1220	2600	0	130	70	160
85	60	0	800	700	700	1220	2600	0	130	70	160
90	60	0	800	700	700	1220	2600	0	130	70	160
95	60	0	750	663	650	1195	2575	0	130	70	160
100	60	0	700	625	600	1170	2550	0	130	70	160
105	60	0	650	588	550	1145	2525	0	130	70	160
110	60	0	600	550	500	1120	2500	0	130	70	160
115	60	0	550	513	450	1095	2475	0	130	70	160
120	60	0	500	475	400	1070	2450	0	130	70	160
125	60	0	450	438	350	1045	2425	0	130	70	160
130	60	0	400	400	300	1020	2400	0	130	70	160
135	0	0	400	400	300	1020	2400	0	130	70	160

(cont.)

/æ/ series

FRAME	AV	AF	High-high F1	Low-high F1	Low-low F1	F2	F3	AB	B1	B2	B3
0	0	0	200	200	200	1100	2150	0	60	110	130
5	0	0	270	270	270	1177	2203	63	60	110	146
10	0	62	340	340	340	1253	2256	63	60	110	162
15	0	0	410	410	410	1330	2309	63	60	110	177
20	60	0	580	480	480	1406	2362	63	60	110	193
25	60	0	614	514	514	1448	2370	63	62	115	209
30	60	0	648	548	548	1491	2377	63	64	120	225
35	60	0	682	582	582	1533	2385	63	66	125	241
40	60	0	716	616	616	1575	2392	63	68	130	257
45	60	0	750	650	650	1618	2400	0	70	135	272
50	60	0	750	650	650	1660	2410	0	70	140	288
55	60	0	750	650	650	1660	2420	0	70	145	304
60	60	0	750	650	650	1660	2430	0	70	150	320
65	60	0	750	650	650	1660	2430	0	70	150	320
70	60	0	750	650	650	1660	2430	0	70	150	320
75	60	0	750	650	650	1660	2430	0	70	150	320
80	60	0	750	650	650	1660	2430	0	70	150	320
85	60	0	750	650	650	1660	2430	0	70	150	320
90	60	0	750	650	650	1660	2430	0	70	150	320
95	60	0	700	613	600	1635	2405	0	70	150	320
100	60	0	650	575	550	1610	2380	0	70	150	320
105	60	0	600	538	500	1585	2355	0	70	150	320
110	60	0	550	500	450	1560	2330	0	70	150	320
115	60	0	500	463	400	1535	2305	0	70	150	320
120	60	0	450	425	350	1510	2280	0	70	150	320
125	60	0	400	388	300	1485	2255	0	70	150	320
130	60	0	350	350	250	1460	2230	0	70	150	320
135	0	0	350	350	250	1460	2230	0	70	150	320

[RESEARCH ON SPEECH PERCEPTION Progress Report No. 13 (1987) Indiana University]

Comparative Research on Language Learning*

Judith A. Gierut

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405

---

*This research was supported, in part, by a National Institutes of Health Training Grant NS-07134-09 to Indiana University, Bloomington. I would like to thank Kathleen Bardovi-Harlig, Daniel Dinnsen, and Fred Eckman for their valuable comments on an earlier version of this paper. Portions of this paper were presented at the 1986 Annual Meeting of the American Association for Applied Linguistics, New York.

Abstract

This paper integrates the research concerns of two language learning populations, adults acquiring a second language and children learning to correct functional (nonorganic) speech sound errors. Phonology was specifically examined with regard to four areas of mutual concern and benefit: (a) characterization of the sound system, (b) selection of aspects of the target sound system to be taught, (c) projection of learning during instruction, and (d) application of research findings to classroom and clinic. This comparative research indicated that basic theoretical and pedagogical aims are identical for both populations. Also, research on both populations has resulted in similar findings about language, learning, and instruction. Moreover, the study of each population has shown certain advances that may contribute to, and shape the direction of, language learning research for the other population. Integrated research efforts of this type have potential for isolating properties that are necessary and specific to language from those that are unique to acquisition, and further, for differentiating these universal properties from those that are specific to given language learning populations.

Comparative Research on Language Learning

Linguists have been interested in data from language learning populations for at least three reasons: (a) to further their understanding of the nature and structure of language, (b) to gain insight into the process of language acquisition and learning, and (c) to study specific subgroups of language learners.

Perhaps, the most obvious and direct use of language learning data has been to examine particular populations of learners, such as blind or deaf children acquiring language. Focusing on particular populations of language learners provides information about the nature and emergence of the linguistic systems of these speakers. For example, the study of children with functional (nonorganically-based) speech disorders has led to the observation that these children typically do not exhibit "deviant" language systems (Dinnsen, Elbert, & Weismer, 1980; Gandour, 1981; Haas, 1963; Leonard, 1973). Rather, these children display language systems that may be developmentally delayed and/or different from the adult target, but that are generally consistent with properties and features of primary languages. From careful study of this language learning population, then, a priori assumptions about the nature and origin of functional speech disorders have been modified.

A second way in which language learning data have been used is in the formulation and confirmation of linguistic theories (Ferguson, 1975, 1977; Fromkin, 1987; Gandour, 1981; Jakobson, 1941; Shattuck-Hufnagel & Klatt, 1979; Smith, 1973). Studies of language learning populations, such as dyslexic or aphasic adults, may force one to abandon existing methodologies and frameworks; as a result, new insights into the nature of language are often gained. The study of language learning populations from this perspective also contributes important information about those aspects of language that are innate ("acquired") versus those that are learned and, further, those aspects of grammar that are necessary and specific to language versus those that are essential to cognition.

Conclusions drawn from language learning data in and of itself or in support of linguistic theory, however, may be limited. Linguistic skills or learning patterns observed in a given population may be representative of more general features of acquisition or of language and may not be indicative of the unique characteristics of a population. Conversely, information about language and learning may be peculiar to a specific population and may not be generalizable to broader aspects of language or acquisition. Language learning data must be examined in alternate ways in order to factor out properties universal to language from those unique to acquisition, and then, to differentiate these universal properties from those that are specific to a given language learning population. One way this may be accomplished is through comparative research across language learning populations. Mutual benefits, both theoretical and pedagogical, may obtain when the results and methodologies of research on language learning in one population are integrated and shared with those of another population. For example, tracing the course of language acquisition in normally developing children has helped linguists predict certain patterns and sequences of learning in adults acquiring a second language (Dulay, Burt, & Krashen, 1982; Flege & Davidian, 1984; Hecht & Mulford, 1982; Johansson, 1973; Wode, 1981). Aspects of language development common to these two populations may constitute some of the essential or basic elements of language learning. As another example, identifying the locus of perceptual, productive, and processing difficulties in adults with dyslexia and those with Alzheimer's disease has facilitated methods of language rehabilitation for aphasic adults (Lieberman, Meskill, Chatillon, & Shupack, 1985; Nicholas, Obler, Albert, & Helm-Estabrooks, 1985; Rastatter & Lawson-Brill, 1987). Through coordinated research efforts of this type, it should be possible to identify universal versus specific properties of language and the language learning process. To date, however, the relative importance and contribution of comparative and integrated research across language learning populations has not been evaluated.

The purpose of this paper is to examine and integrate the research concerns of two specific language learning populations, adults learning a second language and children learning to overcome functional speech disorders. The specific component of language learning to be examined is phonology. These two populations were selected for comparison because they present no organic or neurological involvement. Both populations also exhibit developing sound systems that are aimed at approximating the target sound system. Moreover, these populations offer a unique testing ground for the study of phonological learning since research in both areas has provided descriptive, instructional, and experimental techniques for the investigation of theoretical and applied questions. Four parallel areas of concern will be examined: (a) characterization of the sound system, (b) selection of certain aspects of the target sound system to be taught, (c) projection of learning during instruction, and (d) application of research findings to classroom and clinic. These research concerns, while not the only areas of overlap, were selected because they represent core components of language learning and instruction (Gierut, 1985b; Gierut & Dinnsen, 1987).

Characterization of the Sound System

The phonological systems of second language learners have been described as independent of both the native and the target language (Bialystok & Sharwood Smith, 1985; Dickerson, 1975; Eckman, 1981b; Selinker, 1969, 1972), hence, the "interlanguage." The disordered sound systems of young children have likewise been described as independent of the target or adult sound system (Camarata & Gandour, 1984, 1985; Dinnsen, 1984; Dinnsen et al., 1980; Fey & Stalker, 1986; Gandour, 1981; Gierut, 1985c; Maxwell, 1981; Williams & Dinnsen, 1987).[1] Thus, both second language learners and speech disordered children maintain unique phonological systems, independent of the target, in terms of both the structure and function of sounds.

The sound systems of these learners also bear structural similarity to each other. The sound systems have been shown to be systematic, characterized by phonological rules, both allophonic and neutralizing, and by phonotactic constraints (Camarata & Gandour, 1984; Dickerson, 1975; Dickerson, 1976; Dinnsen & Maxwell, 1981; Eckman, 1981a, 1981b; Elbert & Gierut, 1986; Fey & Stalker, 1986; Gierut, 1985a, 1985c, 1986b; Tarone, 1978). Although systematic in nature, the sound systems of these learners have been shown to be highly variable (Dickerson, 1975; Dickerson, 1977; Dinnsen & Elbert, 1984; Gierut, 1986a; Tarone, 1978; Williams, 1980). The locus of phonological variation in second language learners has been associated with sociolinguistic factors such as style shifting (Beebe, 1980; Dickerson, 1975; Tarone, 1979, 1983); whereas, the locus of variation in speech disordered children has not yet been identified.

In addition, the sound systems of these language learners bear similarity to the phonologies of primary languages (Dickerson, 1976; Eckman, 1977, 1981b; Gandour, 1981; Gierut, 1985c, 1986b; see however, Adjemian, 1976; Eckman, 1981b). Moreover, changes observed in these developing sound systems over time bear resemblence to historical sound change (Dickerson, 1976; Gierut, 1985c, 1986b). For the most part, the phonological systems of second language learners and speech disordered children exhibit many of the same properties of natural languages.

Both second language learners and speech disordered children, however, exhibit errors in target sound production. Errors may be due to target-like ("correct") underlying representations affected by phonological rules or nontarget-like ("incorrect") underlying representations characterized by phonotactic constraints. For speech disordered children, errors in target sound production have been associated primarily with nontarget-like underlying representations (Dinnsen, 1986a, 1986b); for second language learners, errors generally result from the application of phonological rules (Gierut & Bardovi-Harlig, in preparation; Gierut, Dinnsen, & Bardovi-Harlig, 1987; Hammerly, 1982).

Also, for both populations, accurate target sound productions may be observed for the "wrong" phonological reason. For example, Eckman (personal communication) observed the case of a Spanish speaker learning English who produced the morphophonemic alternations "smooth" [smut] ~ [smuðɚ] "smoother". These productions derived from the underlying form /smud/, affected by phonological rules of word-final devoicing and intervocalic spirantization, respectively. On the surface, this speaker accurately produced the word "smoother," but only as a result of a phonological rule operating on a nontarget-like underlying representation. Similarly, Dinnsen (personal communication) observed a speech disordered child who did not use /ʃ/ and /tʃ/ contrastively. Moreover, a phonological rule operated in this child's system such that /ʃ/ was realized as [tʃ] word-finally. This child produced morphophonemic alternations between "fish" [fɪtʃ] ~ [fɪʃɪŋ] "fishing". These productions derived from the target-like underlying representation /fɪʃ/. Morphophonemic alternations were also noted between "catch" [kætʃ] ~ [kæʃɪŋ] "catching". Here, the correct production of "catch" derived from a nontarget-like underlying representation, /kæʃ/, affected by a phonological rule. In both of these cases, correct productions for the wrong phonological reason resulted from the operation of an allophonic rule on nontarget-like underlying representations (see also Camarata & Gandour, 1984; Williams & Dinnsen, 1987).

Finally, there are several common research issues related to the characterization of disordered and interlanguage phonologies. Researchers have been concerned, for example, with how these unique and independent phonologies derive (Broselow, 1984; Connell, 1982; Elbert, 1984; Ellis, 1982; Felix, 1980; Hecht & Mulford, 1982; Leonard & Brown, 1984; Tarone, 1980). Do developmental processes or universal constraints shape the organization of the sound system? What is the relative contribution of each of these factors? As another example, researchers in both disciplines have been concerned with how to best characterize sound systems (Dinnsen, 1984; Eckman, 1977, 1985; Elbert & Gierut, 1986; Hammerly, 1982; Sah, 1981; Schachter, 1974; Tarone, 1983). What is the best method for obtaining an objective measure of a speaker's internal knowledge? How can we accurately evaluate aspects of the target system that have already been mastered by a given speaker and those that have yet to be learned?

Selection of Aspects of the Sound System to be Taught

A common, although difficult, task for second language learners is the restructuring of allophones in the native phonology as distinct phonemes in the target phonology, that is, a phonemic split (Lado, 1957). The recommended method for affecting a phonemic split is to teach minimal pair contrasts. At present, there are no reported data on the effectiveness of this method in inducing phonemic splits or on the processes that may be involved in acquiring phonemic splits for second language learners (see, however, Pisoni, Aslin, Perey, & Hennessy, 1982, for an experimental laboratory demonstration of a phonemic split at a perceptual level).

For speech disordered children, the problem of inducing a phonemic split has been of concern only recently (Camarata & Gandour, 1984; Gierut, 1986b; Maxwell, 1981; Williams & Dinnsen, 1987). Borrowing teaching techniques from second language instruction, Gierut (1986b) demonstrated that speech disordered children can learn to reassign allophones as phonemes in the target sound system. Moreover, because the course of learning was monitored systematically and longitudinally, four qualitatively and quantitatively distinct stages in the acquisition of a phonemic split were identified. Specifically, the subject of this study produced [f] and [s] in complementary distribution, such that [f] always and only occurred word-initially and [s] always and only occurred postvocalically. Thus, at Stage 1, no phonemic contrast was present and an allophonic rule was used. With treatment, the subject produced [f] and [s] in all word positions, but only for some morphemes; moreover, alternations between [f] and [s] were observed postvocalically for certain morphemes. At Stage 2, then, a phonemic contrast was present for some morphemes, but this contrast was neutralized. At this stage, there was no evidence that the allophonic rule of Stage 1 continued to operate. With further treatment, the subject produced [f] and [s] in all word positions and neither the allophonic rule of Stage 1 nor the neutralization rule of Stage 2 applied; however, production of [f] and [s] still did not extend to all morphemes. Stage 3, therefore, was characterized by a phonemic contrast in all contexts for most morphemes. Finally, the subject produced [f] and [s] in all contexts for all morphemes and no phonological rules were used. Stage 4 represented a successful phonemic split.

These four stages provide a more detailed picture of the emergence of phonemic splits for speech disordered children. At present, comparable stages of change have not been reported for second language learners; it will be necessary to document longitudinally the degree and extent of change for these speakers as well. Through this type of comparative research, a more fine-grained characterization of the nature and course of acquiring phonemic splits will potentially be developed as well as more effective and efficient procedures for affecting phonemic splits.

Projection of Learning During Instruction

There are at least two ways that learning during instruction has been predicted. A first approach relies on universal properties of language to predict learning; a second approach relies on properties internal to individual speakers.

Language-general Factors

Typological or implicational markedness is one language-general property that has been examined for predictive power in both areas of research, second language acquisition and speech disorders. Eckman and colleagues (Eckman, 1977, 1981a, 1985; Eckman, Moravcsik, & Wirth, 1983, 1985) observed that second language learners who evidenced more marked sounds and sound sequences in the interlanguage also evidenced unmarked sounds and sequences, but not the reverse. Consequently, markedness was suggested as a metric of the degree of difficulty that a second language speaker may have in learning certain target sounds (Eckman, 1977, 1981a, 1985). From a pedagogical point of view, it may be that second language learners who are instructed on more marked errored target segments will spontaneously acquire other related, unmarked target segments that are not directly taught (for comparable examples in interlanguage syntax, see Eckman, 1985, Gass, 1982, and Schachter, 1974).

A similar set of observations has been noted in the area of speech disorders. Dinnsen and Elbert (1984) and Elbert, Dinnsen, and Powell (1984) observed that, during clinical treatment, a child's performance on unmarked errored target sounds was better than his or her performance on marked errored target sounds. These researchers noted, however, that treatment of marked target sounds seemed to result in the acquisition of both marked and unmarked targets (see McReynolds & Jetzke, 1986, for a related observation in remediation of hearing-impaired children).

Descriptive evidence from both disciplines, thus, suggests that the language-general factor of typological markedness may be used to predict learning. It is hypothesized that marked target sounds produced in error may be more difficult to learn initially, but that instruction on these sounds will result in more extensive learning. To date, there has been no experimental evaluation of this hypothesis within either field; this remains a key question for future investigation.

Speaker-specific Factors

A child's competence, or tacit phonological knowledge, of the target sound system is one speaker-specific factor that has been examined in the area of speech disorders. Gierut and colleagues (Gierut, 1985c; Gierut & Dinnsen, 1987; Gierut, Elbert, & Dinnsen, 1987) have experimentally evaluated a child's phonological knowledge as a predictor of learning. Greater amounts of learning were observed in those cases where a child internalized target underlying representations. That is, if a child mastered target underlying representations, even though phonological rules may have been operating to produce errors, performance on these target sounds was better than those cases where target underlying representations had not yet been learned. However, more extensive changes in the overall phonological system were observed when a child was first taught to produce sounds that were most unlike the target language (errored) in terms of underlying representations.

In the area of second language learning, a speaker's phonological knowledge has likewise been cited as a factor that may influence learning (Bialystok, 1981; Bialystok & Sharwood Smith, 1985; Briere, 1966; Dulay et al., 1982; Hammerly, 1982; McLaughlin, 1978). Hammerly (1982) noted that second language learners seem to have most difficulty learning allophonic problems, or cases where target underlying representations have already been internalized. Allophonic problems seemed to be more resistive to change than phonemic problems, or cases where target language underlying representations have not been internalized. This observation is just opposite of that noted for speech disordered children. On the other hand, Briere (1966) demonstrated experimentally that target sounds present in a second language learner's inventory, whether at an underlying or a phonetic level, were learned more rapidly than those target sounds absent from the inventory. This experimental finding is consistent with that reported for speech disordered children. It will be necessary to evaluate experimentally and descriptively these discrepancies between phonological knowledge and learning in speakers acquiring a second language. Pedagogically, it will also be important to examine differences in the amount and extent of learning by second language speakers when instruction begins with target underlying representations that have not yet been learned versus those that have already been mastered.

Application of Research Findings to Classroom and Clinic

In a recent publication, Lightbown (1985) cautioned the direct classroom application of research results. She noted, as have Tarone and others (Tarone, Swain, & Fathman, 1976), that there are several reasons why the classroom application of research findings may be premature. These include, for example, the lack of data on individual learning strategies and styles, the limited information on individual and environmental variables, the generally undeveloped methodology for experimental instructional studies, and the limited number of replications that have been reported. There appears, then, to be a gap between applied research and classroom application in the area of second language instruction. This gap may be partially due to the focus on groups of learners, rather than individuals, in both research and instructional settings. Group research tends to mask important individual differences in learning. The critical assumption is that second language learners are homogeneous and that interlanguage systems are shared by all learners. Moreover, research methodologies involving large numbers of matched subjects often prohibit longitudinal traces of learning or systematic replications of results.

In the area of speech disorders, there has also been somewhat of a dichotomy between the researcher and the clinician. In speech disorders, however, one frequently used experimental paradigm combines the interests of both researcher and clinician, thereby, narrowing this gap between research and application. This experimental paradigm is known as applied behavior analysis, also called functional analysis or single-subject methodology (Hersen & Barlow, 1976; McReynolds & Kearns, 1983).

Single-subject methodology has been widely used in applied disciplines interested in changing a learner's performance through instruction. The logic underlying single-subject methodology is that each subject serves as his or her own control. That is, control over extraneous or interfering variables is demonstrated with the individual subject. Comparisons are made between a subject's performance during periods of no training and training, or no instruction and instruction. The basic assumption is that a subject's performance will not change until instruction is introduced. This assumption is identical to that of other experimental paradigms that employ larger groups of subjects.

There are two essential components of single-subject designs, a no training phase and a training phase. The no training, or baseline, phase serves as a measure of a subject's performance prior to the introduction of training. It is essential that a subject's performance during baseline remains stable in order to demonstrate that training is, in fact, what causes changes in performance. A subject's performance, therefore, must be measured repeatedly during baseline to ensure adequate control. Performance continues to be monitored frequently during training to evaluate degree of learning and training effectiveness. The no training/training phases can be combined or sequenced in a variety of ways across time, behaviors, subjects, or settings.

Single-subject methodology offers several advantages. One advantage is that these designs avoid the problem of identifying large numbers of homogeneous matched subjects. Another advantage of this methodology is that it is possible to look at variation in performance for a given subject as well as across subjects. Sources controlling intra- and intersubject variation can, thus, be identified. A third advantage of this methodology is that behaviors are measured frequently so improvements in performance can be monitored systematically and longitudinally. This provides for an examination of both the spontaneous acquisition of new responses as well as the generalization of learned or treated responses. From these data, an individual subject's learning strategy and style can be determined. Single-subject designs also offer the advantage of being able to test and evaluate different instructional procedures. Finally, single-subject designs are flexible and can be modified in ways directly related to the applied research question (Connell & Thompson, 1986; Kearns, 1986; McReynolds & Thompson, 1986).

One misconception about single-subject research relates to external validity or the generality of research findings. It may be thought that, since single-subject designs do not rely on assumptions of random sampling, this type of research does not generalize from the individual back to the population. This is false; external validity in single-subject research is demonstrated by direct and/or systematic replication of the training effect (McReynolds & Thompson, 1986).

Single-subject research is particularly well-suited to the study of speech disordered children since large numbers of homogeneous and identically matched subjects, necessary for group investigations, are generally unavailable. Also, this research methodology closely parallels the typical clinical training situation, namely, one-on-one instruction. Moreover, clinicians and researchers alike are interested in developing effective training programs supported by experimental data, and in using these data to come to a more basic understanding of language and the language learning process.

Single-subject design may likewise help bridge the gap between research and application in second language instruction, providing a well-developed, sophisticated methodology for evaluating instructional techniques, for determining individual learning strategies, and for establishing the role and function of social and environmental factors. That interlanguages may not be universally shared among second language learners (Bialystok & Sharwood Smith, 1985; Eckman, 1981a, 1985) and that second language learners may constitute a heterogeneous group (Gierut & Bardovi-Harlig, in preparation) further supports the importance of using single-subject rather than group designs. Emphasis on the individual in single-subject research, however, may necessitate certain modifications in the classroom approach to second language instruction. The nature of such changes will depend upon the results of experimental studies that examine factors affecting a given speaker's phonological learning. It remains for future investigation to determine whether instruction should focus on areas of phonological difficulty common to second language speakers or whether emphasis should be placed on individualized areas of difficulty.

Conclusion

The comparative and integrated approach to language learning set forth in this paper has contributed specifically to our understanding of the nature and interaction among two particular subgroups of learners, adults acquiring a second language and children learning to correct speech errors. From this comparison, it has been demonstrated that (a) the study of these two populations is based on similar research and pedagogical aims, (b) research in both disciplines has led to similar findings about language, learning, and instruction, and (c) each discipline shows certain advances in different aspects of language learning research that may benefit the other, theoretically and pedagogically.

The present comparison is limited, however, in that it focuses on only two language learning groups. The converging findings make a preliminary contribution to the identification of universal properties of language and language learning. It will be important to examine other subgroups of learners on the same points of research concern in order to fully differentiate among language learning populations and to glean generalities about the nature of language and acquisition.

This comparison of second language learners and speech disordered children, thus, has been a first attempt at illustrating the importance and potential contribution of integrated research efforts; potentially, it will also serve as an impetus for continued research of this type. Comparative research lends itself well to furthering our understanding of specific language learning populations, to improving the effectiveness of our instructional methods, and to identifying those properties necessary and specific to language and the language learning process.

Endnotes

1 There are several approaches to the analysis and characterization of speech sound disorders (see Elbert & Gierut, 1986, for review). These include, for example, place-voice-manner analysis, standard generative analysis, and natural process analysis. With exception of generative analysis, these approaches assume that a child's knowledge of the sound system is identical to that of the adult's at an underlying level; however, at a surface phonetic level, a child's knowledge of the sound system may be different than the adult's. It has been argued that this assumption is neither necessary nor sufficient (Camarata & Gandour, 1984; Dinnsen, 1984; Maxwell, 1981, 1984; Williams & Dinnsen, 1987); thus, claims about a child's phonological system being independent of the adult target are based upon generative phonological descriptions.

References

Adjemian, C. (1976). On the nature of interlanguage systems. *Language Learning*, *26*, 297-320.

Beebe, L. (1980). Sociolinguistic variation and style shifting in second language acquisition. *Language Learning*, *30*, 443-447.

Bialystok, E. (1981). The role of linguistic knowledge in second languge use. *Studies in Second Language Acquisition*, *4*, 31-45.

Bialystok, E., & Sharwood Smith, M. (1985). Interlanguage is not a state of mind: An evaluation of the construct for second language acquisition. *Applied Linguistics*, *6*, 101-117.

Briere, E.J. (1966). An investigation of phonological interference. *Language*, *42*, 768-796.

Broselow, E. (1984). An investigation of transfer in second language phonology. *International Review of Applied Linguistics*, *22*, 253-269.

Camarata, S., & Gandour, J. (1984). On describing idiosyncratic phonologic systems. *Journal of Speech and Hearing Disorders*, *49*, 262-266.

Camarata, S., & Gandour, J. (1985). Rule invention in the acquisition of morphology by a language-impaired child. *Journal of Speech and Hearing Disorders*, *50*, 40-45.

Connell, P. (1982, November). *Markedness differences in the substitutions of normal and misarticulating children*. Paper presented at the Annual Convention of the American Speech, Language, and Hearing Association, Toronto.

Connell, P., & Thompson, C.K. (1986). Flexibility of single-subject experimental designs. Part III: Using flexibility to design or modify experiments. *Journal of Speech and Hearing Disorders*, *51*, 214-225.

Dickerson, L.J. (1975). The learner's interlanguage as a system of variable rules. *TESOL Quarterly*, *9*, 401-407.

Dickerson, W.B. (1976). The psycholinguistic unity of language learning and language change. *Language Learning*, *26*, 215-231.

Dickerson, W.B. (1977). Language variation in applied linguistics. *ITL Review of Applied Linguistics*, *35*, 43-66.

Dinnsen, D.A. (1984). Methods and empirical issues in analyzing functional misarticulations. In M. Elbert, D.A. Dinnsen, & G. Weismer (Eds.), *Phonological theory and the misarticulating child* (ASHA Monograph No. 22, pp. 5-17). Rockville, MD: ASHA.

Dinnsen, D.A. (1986a, November). *Fundamental characteristics of disordered phonological systems*. Paper presented at the Annual Convention of the American Speech, Language, and Hearing Association, Detroit.

Dinnsen, D.A. (1986b, November). *On the explanation of changes in phonological knowledge*. Paper presented at the Annual Convention of the American Speech, Language, and Hearing Association, Detroit.

Dinnsen, D.A., & Elbert, M. (1984). On the relationship between phonology and learning. In M. Elbert, D.A. Dinnsen, & G. Weismer (Eds.), *Phonological theory and the misarticulating child* (ASHA Monograph No. 22, pp. 59-68). Rockville, MD: ASHA.

Dinnsen, D.A., Elbert, M., & Weismer, G. (1980). Some typological properties of functional misarticulation systems. In W.O. Dressler (Ed.), *Phonologica 1980* (pp. 83-88). Innsbruck: Innsbrucker Beitrage Zur Sprachwissenschaft.

Dinnsen, D.A., & Maxwell, E.M. (1981). Some phonology problems from functional speech disorders. *Innovations in Linguistic Education*, *2*, 79-98.

Dulay, H., Burt, M., & Krashen, S. (1982). *Language two*. New York: Oxford University Press.

Eckman, F. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, *27*, 315-330.

Eckman, F. (1981a). On predicting phonological difficulty in second language acquisition. *Studies in Second Language Acquisition*, *4*, 18-30.

Eckman, F. (1981b). On the naturalness of interlanguage phonological rules. *Language Learning*, *31*, 195-216.

Eckman, F. (1985). Some theoretical and pedagogical implications of the markedness differential hypothesis. *Studies in Second Language Acquisition*, *7*, 289-307.

Eckman, F., Moravcsik, E., & Wirth, J. (1983, December). *On interlanguage and language universals*. Paper presented at the Winter Meeting of the Linguistics Society of America, Minneapolis.

Eckman, F., Moravcsik, E., & Wirth, J. (1985, October). *Language learning and language typology*. Paper presented at the Midwest TESOL Regional Conference, Milwaukee.

Elbert, M. (1984). The relationship between normal phonological acquisition and clinical intervention. In N.J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 10, pp. 111-139). New York: Academic Press.

Elbert, M., Dinnsen, D.A., & Powell, T.W. (1984). On the prediction of phonologic generalization learning patterns. *Journal of Speech and Hearing Disorders*, *49*, 309-317.

Elbert, M., & Gierut, J.A. (1986). *Handbook of clinical phonology: Approaches to assessment and intervention*. San Diego: College-Hill Press.

Elbert, M., & McReynolds, L.V. (1979). Aspects of phonological acquisition during articulation training. *Journal of Speech and Hearing Disorders*, *44*, 459-471.

Ellis, R. (1982). The origins of interlanguage. *Applied Linguistics*, *3*, 207-223.

Felix, S. (1980). Interference, interlanguage, and related issues. In S. Felix (Ed.), *Second langue development: Trends and issues* (pp. 93-107). Tubingen: Gunter Narr Verlag.

Ferguson, C.A. (1975). Sound patterns in language acquisition. In D.P. Dato (Ed.), *Georgetown University round table on language and linguistics 1975* (pp. 1-16). Washington, D.C.: Georgetown University Press.

Ferguson, C.A. (1977). New directions in phonological theory: Language acquisition and universals research. In R.W. Cole (Ed.), *Current issues in linguistic theory* (pp. 247-299). Bloomington, IN: Indiana University Press.

Fey, M., & Stalker, C. (1986). A hypothesis-testing approach to treatment of a child with an idiosyncratic (morpho)phonological system. *Journal of Speech and Hearing Disorders*, *51*, 324-336.

Flege, J.E., & Davidian, R. (1984). Transfer and developmental processes in adult foreign language speech production. *Applied Psycholinguistics*, *5*, 323-347.

Fromkin, V.A. (1987). The lexicon. *Language*, *63*, 1-22.

Gandour, J. (1981). The nondeviant nature of deviant phonological systems. *Journal of Communication Disorders*, *14*, 11-29.

Gass, S. (1982). From theory to practice. In L. Crymes & W. Rutherford (Eds.), *On TESOL 82*. Washington, D.C.: TESOL.

Gierut, J.A. (1985a). Generative phonology: Clinical applications in speech pathology. *Innovations in Linguistic Education*, *3*, 152-167.

Gierut, J.A. (1985b, December). *On predicting generalization in phonological learning*. Paper presented at the Annual Meeting of the American Association for Applied Linguistics, Seattle.

Gierut, J.A. (1985c). *On the relationship between phonological knowledge and generalization learning in misarticulating children*. Doctoral dissertation, Indiana University, Bloomington. (Also distributed by the Indiana University Linguistics Club, 1986).

Gierut, J.A. (1986a, November). *On characterizing variability in phonologically disordered speech*. Paper presented at the Annual Convention of the American Speech, Language, and Hearing Association, Detroit.

Gierut, J.A. (1986b). Sound change: A phonemic split in a misarticulating child. *Applied Psycholinguistics*, *7*, 57-68.

Gierut, J.A., & Bardovi-Harlig, K. (in preparation). Individual differences in interlanguage phonology.

Gierut, J.A., & Dinnsen, D.A. (1987). On predicting ease of phonological learning. *Applied Linguistics*, *8*, 35-57.

Gierut, J.A., Dinnsen, D.A., & Bardovi-Harlig, K. (1987). External validity of productive phonological knowledge: A first report. *Research on speech perception progress report no. 13*. Bloomington, IN: Speech Research Laboratory, Department of Psychology, Indiana University.

Gierut, J.A., Elbert, M., & Dinnsen, D.A. (1987). A functional analysis of phonological knowledge and generalization learning in misarticulating children. *Journal of Speech and Hearing Research*, *30*, 462-479.

Haas, W. (1963). Phonological analysis of a case of dyslalia. *Journal of Speech and Hearing Disorders*, *28*, 239-246.

Hammerly, H. (1982). Contrastive phonology and error analysis. *International Review of Applied Linguistics*, *20*, 17-32.

Hecht, B.F., & Mulford, R. (1982). The acquisition of a second language phonology: Interaction of transfer and developmental factors. *Applied Psycholinguistics*, *3*, 313-328.

Hersen, M., & Barlow, D.H. (1976). *Single case experimental designs: Strategies for studying behavior change*. New York: Pergamon Press.

Jakobson, R. (1941). *Child language, aphasia, and phonological universals* (A.R. Keiler, Trans.). The Hague: Mouton.

Johansson, F.A. (1973). *Immigrant Swedish phonology: A study of multiple contact analysis*. Lund, Sweden: CWK Gleerup.

Kearns, K.P. (1986). Flexibility of single-subject experimental designs. Part II: Design selection and arrangement of experimental phases. *Journal of Speech and Hearing Disorders*, *51*, 204-214.

Lado, R. (1957). *Linguistics across cultures*. Ann Arbor, MI: The University of Michigan Press.

Leonard, L.B. (1973). The nature of deviant articulation. *Journal of Speech and Hearing Disorders*, *38*, 156-161.

Leonard, L.B., & Brown, B.L. (1984). Nature and boundaries of phonologic categories: A case study of an unusual phonologic pattern in a language-impaired child. *Journal of Speech and Hearing Disorders*, *49*, 419-428.

Lieberman, P., Meskill, R.H., Chatillon, M., & Shupack, H. (1985). Phonetic speech perception deficits in dyslexia. *Journal of Speech and Hearing Research*, *28*, 480-486.

Lightbown, P. (1985). Great expectations: Second language acquisition research and classroom teaching. *Applied Linguistics*, *6*, 173-189.

Maxwell, E. M. (1981). A study of misarticulation from a linguistic perspective. Doctoral dissertation, Indiana University, Bloomington. (Also distributed by the Indiana University Linguistics Club, 1982).

Maxwell, E.M. (1984). On determining underlying phonological representations of children: A critique of the current theories. In M. Elbert, D.A. Dinnsen, & G. Weismer (Eds.), Phonological theory and the misarticulating child (ASHA Monographs No. 22, pp. 18-29). Rockville, MD: ASHA.

McLaughlin, B. (1978). The monitor model: Some methodological considerations. Language Learning, 28, 309-332.

McReynolds, L.V., & Jetzke, E. (1986). Articulation generalization of voiced-voiceless sounds in hearing-impaired children. Journal of Speech and Hearing Disorders, 51, 348-355.

McReynolds, L.V., & Kearns, K.P. (1983). Single-subject experimental designs in communicative disorders. Baltimore: University Park Press.

McReynolds, L.V., & Thompson, C. (1986). Flexibility of single-subject experimental designs. Part I: Review of the basics of single-subject designs. Journal of Speech and Hearing Disorders, 51, 194-203.

Nicholas, M., Obler, L.K., Albert, M.L., & Helm-Estabrooks, N. (1985). Empty speech in Alzheimer's disease and fluent aphasia. Journal of Speech and Hearing Research, 28, 405-410.

Pisoni, D.B., Aslin, R.N., Perey, A.J., & Hennessy, B.L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. Journal of Experimental Psychology: Human Perception and Performance, 8, 297-314.

Rastatter, M.P., & Lawson-Brill, C. (1987). Reaction times of aging subjects to monaural verbal stimuli: Some evidence for a reduction in right-hemisphere linguistic processing capacity. Journal of Speech and Hearing Research, 30, 261-267.

Sah, P. (1981). Contrastive analysis, error analysis, and transformational generative theory: Some methodological issues in the theory of second language learning. International Review of Applied Linguistics, 19, 95-112.

Schachter, J. (1974). An error in error analysis. Language Learning, 24, 205-214.

Selinker, L. (1969). Language transfer. General Linguistics, 9, 67-92.

Selinker, L. (1972). Interlanguage. International Review of Applied Linguistics, 10, 209-231.

Shattuck-Hufnagel, S., & Klatt, D.H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. Journal of Verbal Learning and Verbal Behavior, 18, 41-55.

Smith, N.V. (1973). *The acquisition of phonology: A case study*. Cambridge: Cambridge University Press.

Tarone, E. (1978). The phonology of interlanguage. In J. Richards (Ed.), *Understanding second and foreign language learning* (pp. 15-33). Rowley, MA: Newbury House.

Tarone, E. (1979). Interlanguage as chameleon. *Language Learning*, *29*, 181-191.

Tarone, E. (1980). Some influences on the syllable structure of interlanguage phonology. *International Review of Applied Linguistics*, *18*, 139-152.

Tarone, E. (1983). On the variability of interlanguage systems. *Applied Linguistics*, *4*, 142-164.

Tarone, E., Swain, M., & Fathman, A. (1976). Some limitations to the classroom applications of current second language acquisition research. *TESOL Quarterly*, *10*, 19-31.

Williams, A.L., & Dinnsen, D.A. (1987). A problem of allophonic variation in a speech disordered child. *Innovations in Linguistic Education*, *5*, 85-90.

Williams, L. (1980). Phonetic variation as a function of second-language learning. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child phonology: Perception* (Vol. 2, pp. 185-216). New York: Academic Press.

Wode, H. (1981). *Learning a second language, I: An integrated view of language acquisition*. Tubingen: Gunter Narr Verlag.

[RESEARCH ON SPEECH PERCEPTION Progress Report No. 13 (1987) Indiana University]

Maximal Opposition Approach to Phonological Treatment*

Judith A. Gierut

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, IN 47405

*This research was supported, in part, by a National Institutes of Health Training Grant (NS-07134-09) to Indiana University, Bloomington. I would like to thank Steve Chin, Gladys DeVane, and Karen Hardin for their assistance with various aspects of clinical intervention and interjudge reliability. Dan Dinnsen, Mary Louise Edwards, and Marc Fey provided helpful comments on an earlier version of this paper. Portions of this paper were presented at the 1987 American Speech-Language-Hearing Association Convention, New Orleans.

Abstract

The purpose of this paper was to evaluate a phonological treatment program of maximal rather than minimal feature contrasts by charting the course of learning in a child displaying a systematic error pattern involving the nonoccurrence of word-initial consonants. Generalization data indicated that the child learned 16 word-initial consonants following treatment of only 3 sets of maximal opposition contrasts. Overgeneralization data indicated that the child restructured his phonological system based on a larger concept of "word-initialness." Basic components of, and differences between various forms of contrast treatment are discussed.

Maximal Opposition Approach to Phonological Treatment

Minimal pair contrast treatment is one method of remediation that has been used to improve and change the phonological systems of children displaying speech sound errors (Elbert, Rockman, & Saltzman, 1980; Ferrier & Davis, 1973; Weiner, 1981; Winitz, 1975). Minimal pair contrast treatment typically involves having a child distinguish - through discrimination, imitation, and/or spontaneous production - pairs of syllables or words that are unique along a single feature or dimension.[1] For example, the word pairs "pig"-"big," "tip"-"dip," and "coat"-"goat" are each minimally contrastive in terms of voicing word-initially. The voicing distinction makes these word-initial sounds phonemic in English and, consequently, these word pairs are lexically unique. Through minimal pair contrast treatment, a child is taught that different sounds signal different meanings. Minimal pair treatment thus enhances a child's conception of sounds as phonemes (Weiner, 1981). Moreover, minimal pair contrast treatment reduces the occurrence of homonomy in a child's productions by contrasting desired target sounds with error or substituted sounds (Ingram, 1976).

Minimal pair contrast treatment has been widely employed by clinical researchers adopting a variety of assessment-intervention frameworks including, among others, a distinctive feature approach (e.g., Costello & Onstine, 1976; McReynolds & Bennett, 1972), a phonological process approach (e.g., Weber, 1970; Weiner, 1981), and a standard generative approach (Elbert, Dinnsen, & Powell, 1984; Gierut, Elbert, & Dinnsen, 1987). This form of treatment has been successful in facilitating the acquisition of specific and trained minimal pairs, as well as in enhancing generalization of other sound or word pairs that vary along similar dimensions. For example, within a phonological process framework, Weiner (1981) reduced the frequency of final consonant deletion, stopping, and velar fronting by teaching children meaningful minimal pairs such as "pie"-"pipe," "see"-"tea," and "gate"-"date," respectively. Weiner further observed that the use of these phonological processes was reduced in other untreated words following contrast treatment. Similarly, within a distinctive feature framework, McReynolds and Bennett (1972) taught a child the contrast between /ʃ/ and /tʃ/, differing in the continuancy feature. Following treatment, generalization to other continuant sounds (i.e., /f,v,s,z/) was noted. Thus, focusing a child's attention specifically on a single feature that uniquely distinguishes one sound or word from another appears to result in both the learning and generalization of aspects of phonology.

Recently, an alternate form of contrast treatment has been introduced that also may be clinically relevant (Elbert & Gierut, 1986). This form of contrast treatment involves maximal rather than minimal oppositions. In this approach, phonemic distinctions vary along extremes of the broad and multiple dimensions of voice, place, and manner. Some examples of maximally opposed distinctions include the contrast between a voiced bilabial sonorant /m/ and voiceless velar obstruent /k/ or the contrast between a voiced bilabial stop /b/ and voiceless palato-alveolar fricative /ʃ/. Notice that contrastive sounds for treatment are maximally distinct along several feature dimensions, as compared to minimal pair treatment where phonemic distinctions vary along narrow, binary dimensions such as voiced versus voiceless. The rationale behind a maximal opposition approach to contrast treatment is to provide a child with an opportunity to learn about the target phonology in his or her own unique way by filling in gaps along these extremes of multiple feature dimensions. Presumably, treatment of maximal distinctions allows a child to choose and to attend to those specific feature dimensions that he or she identifies as relevant to sound production. Potentially, the child will focus on target sounds that maintain these relevant distinctions and will generalize accurate production to these particular sounds.

Current literature in developmental psycholinguistics and cognitive psychology supports a maximal opposition approach to phonological treatment. Specifically, a method of maximal oppositions is consistent with the work of Jakobson (1941/1968) and others (Crocker, 1969; Leopold, 1947; Velten, 1943). Young normally developing children initially seem to attempt and to maintain maximal distinctions and contrasts among sounds and sound classes. With development and experience, sound contrasts progress from major oppositions, such as oral-nasal or obstruent-sonorant, to more finely differentiated distinctions varying along the multiple dimensions of voice, place, and manner. These observations suggest that children may first concentrate on the wide extremes of sound contrasts, rather than on fine-grained minimal distinctions.

Other research in developmental psycholinguistics indicates that young children actively participate in the process of phonological acquisition. Evidence has shown that children individually and uniquely select the type of sounds and contrasts that are added to their phonological systems (Ferguson & Farwell, 1975; Ferguson, Peizer, & Weeks, 1973; Menn, 1976; Schwartz & Leonard, 1982; Vihman, 1981). Moreover, children initiate and invent creative solutions to the "puzzle" of phonological acquisition (Ferguson & Macken, 1980; Fey & Gandour, 1982; Macken & Ferguson, 1983; Priestly, 1977). The child is afforded and uses many degrees of freedom in the acquisition of phonology. Perhaps, children with phonological disorders may also benefit from active, creative participation in selecting or changing elements and contrastive aspects of their phonologies in the course of clinical intervention (Elbert, 1984; Fey & Stalker, 1986; cf. Bates, 1976; Bates & MacWhinney, 1982; Johnston, 1982 in the acquisition of syntax).

Finally, research in the area of generalization has suggested that the transfer of learning may be enhanced and facilitated by a "loosely structured" intervention plan (Stokes & Baer, 1977; Leonard, 1981). Ideally, a loosely structured plan does not narrowly limit the treatment items or stimuli used, nor does it restrict the range of correct responses that are allowed. Loosely structured intervention presumably permits a child to sample relevant dimensions of varied treatment items for transfer to other new items and situations, thereby, resulting in widespread generalization.

These three areas of research thus motivate a maximal opposition approach to phonological treatment that (a) emphasizes phonemic contrasts along a more grossly differentiated range of features, (b) allows a child considerable flexibility in identification of relevant feature contrasts, and (c) encourages broad generalization of those features identified as relevant. The purpose of this paper is to describe and evaluate such a treatment approach of maximal opposition. The effectiveness of this program will be evaluated by tracing patterns of phonological generalization and overgeneralization in a child displaying a systematic sound pattern involving the nonoccurrence of consonants word-initially.

Subject

The subject of this study was a boy, J, age 4 years, 7 months. J was referred to the Speech and Hearing Clinic at Indiana University for a diagnostic evaluation at age 4 years, 1 month by his mother due to the unintelligibility of his speech. J displayed numerous sound errors in conversational speech as well as in performance on the Goldman-Fristoe Test of Articulation (Goldman & Fristoe, 1969). Errors were characterized primarily by the nonoccurrence of word-initial consonants. Results of the diagnostic intake indicated that J had normal hearing bilaterally with no history of middle ear infections. Also, J's performance on the Preschool Language Scale-Revised (Zimmerman, Steiner, & Pond, 1979) was age-appropriate both receptively (point score=33; age equivalency=5 years) and expressively (point score=27; age equivalency=4 years, 4.5 months). Parental report indicated that J had no apparent gross or fine motor, cognitive, social, or emotional disfunctions. J's history, however, revealed a secondary cleft of the hard and soft palates which was surgically repaired two years prior to the diagnostic evaluation at the Speech and Hearing Clinic. The physician's report stated that no further medical or dental procedures were necessary and that the child sustained adequate velopharyngeal closure for speech production; an examination of the child's oral mechanism by the speech-language diagnostician corroborated the latter observation. There would appear to be no necessary connection between this child's word-initial omissions and his history of secondary cleft palate since research has shown that children with a history of cleft palate typically exhibit more errors in medial position than initial position and more errors of substitution than omission (Philips & Harrison, 1969). Moreover, J did not evidence other patterns characteristic of cleft palate speech, such as excessive nasality, nasal emission, or snorting. J was from a monolingual English-speaking family.

Phonological Description

Analysis Procedures

A standard generative phonological description of this child's speech was developed prior to treatment using procedures outlined by Dinnsen (1984), Elbert and Gierut (1986), and Gierut (1986). That is, spontaneous connected speech and citation form samples were obtained using story-telling and picture-naming tasks. The citation form sample provided the child with an opportunity to produce all target English sounds in each relevant position (initial, intervocalic, and final) in a minimum of five different exemplars (Gierut, 1985). The citation form sample also provided for the elicitation of potential minimal pairs (e.g., "pig" - "big") and morphophonemic alternations (e.g., "pig" - "piggie"). Speech samples were tape-recorded and then narrowly transcribed using standard notation of the International Phonetic Alphabet. These speech samples served as the data base for developing the generative analysis of J's sound system.

Analysis Results

J's phonetic inventory included production of all target English sounds, with the exception of [f,v,r]. Production of [f,v] was restricted by an inventory constraint; that is, [f,v] never occurred in any word position. Target /r/ was distorted. The sounds present in J's phonetic inventory were used consistently and contrastively as phonemes, but only in postvocalic positions. Word-initially, only a limited subset of phonemes were used, namely, /m,b,w,j/. J produced the majority of morphemes without word-initial consonants. For the most part, word-initial position was not marked productively by the presence of consonants.

To determine more fully whether J had productive knowledge of word-initial sounds, additional morphophonemic data were obtained following a procedure described by Rockman, Dinnsen, and Rowland (1983) and Gierut (1985). These data took the form of adding the prefix "re-" to target morphemes with word-initial consonants and glides. J was instructed to create nonsense words by saying "re-" before the name of pictured stimulus items consisting of word-initial target sounds. The addition of a prefix to a morpheme serves the purpose of altering the phonetic environment, in this case, from word-initial position to postvocalic position (e.g., "cut"-"recut" or "jump"-"rejump"). Given that J produced target sounds postvocalically, but not initially, we might expect that prefixed forms would be produced with morpheme-initial consonants, but nonprefixed forms would be produced without morpheme-initial consonants, as in the examples, "cut" [ʌt] ~ [rikʌt] "recut" or "jump" [ʌmp] ~ [ridʒʌmp] "rejump." Evidence of this type would suggest that J lexically (i.e., underlyingly) represented morphemes for purposes of production with target-appropriate word-initial consonants. J's repertoire of word-initial consonants thus would be relatively complete; however, a phonological rule would be motivated to delete certain word-initial consonants. On the other hand, it was entirely possible that J would produce both prefixed and nonprefixed forms without morpheme-initial consonants, as in the examples "cut" [ʌt] ~ [riʌt] "recut" or "jump" [ʌmp] ~ [riʌmp] "rejump." Evidence of this type would suggest that J lexically represented morphemes for production purposes without target-appropriate word-initial consonants. In this case, J's repertoire of word-initial consonants would be severely limited. A positional constraint would be motivated to exclude certain consonants from word-initial position both lexically and phonetically.

Elicitation of prefixed citation form items (see appendix) suggested that the latter hypothesis was correct. Both prefixed and nonprefixed forms were produced without morpheme-initial consonants, as in the examples shown in Table 1. As expected from this child's pattern of production, however, target /m,b,w,j/ morphemes were marked word-initially whether or not the prefix was added. Thus, it appeared that J did not represent most morphemes lexically or phonetically with word-initial consonants.

Insert Table 1 about here

It is, of course, possible that the prefix procedure was not sensitive enough to induce "true" morphophonemic alternations due to the specific nature of English morphology. In English, suffixes are both derivational (e.g., "swift**ly**," "neat**ness**") and inflectional (e.g., "laugh**ing**," "walk**ed**"); however, prefixes are only derivational. Consequently, a word boundary is maintained between a prefix and the base morpheme to which it is added, technically leaving consonants of the base morpheme in word-initial position. If J had produced morphophonemic alternations between prefixed and nonprefixed forms, we could take this as direct evidence that the child's phonology included word-initial consonants. However, in this case, the absence of morphophonemic alternations does not provide absolute or conclusive evidence about J's

Table 1

Examples of J's Production of Nonprefixed and Prefixed Morphemes

Target /k/	("re+")	"cup"	[ʌp]	~	[w^{r}ɪʌp]
		"cut"	[ʌt]	~	[w^{r}ɪʌt]
		"coat"	[out]	~	[w^{r}iout]
		"comb"	[oum]	~	[w^{r}ioum]
Target /s/	("re+")	"soup"	[up]	~	[w^{r}iup]
		"soap"	[oup]	~	[w^{r}ioup]
		"sock"	[aks]	~	[w^{r}iaks]
		"sun"	[ʌn]	~	[w^{r}ɪʌn]
		"santa"	[ænə]	~	[w^{r}iænə]
Target /tʃ/	("re+")	"chair"	[ɛw^{r}]	~	[w^{r}iɛw^{r}]
		"cheese"	[iz]	~	[w^{r}ɪiz]
		"chip"	[ɪp]	~	[w^{r}iɪp]
Target /m/	("re+")	"mud"	[mʌb]	~	[w^{r}imʌb]
		"mouth"	[mau]	~	[w^{r}imau]
		"mouse"	[maus]	~	[w^{r}imaus]
		"moon"	[mum]	~	[w^{r}imum]
		"mother"	[mamə]		

Table 1 (cont.)

Target /b/	("re+")	"big"	[bɪg]	~	[wʳibɪg]
		"book"	[buk]	~	[wʳibuk]
		"bed"	[bɛb]	~	[wʳibɛb]
		"bus"	[bʌʃ]	~	[wʳibʌs]
		"boot"	[but]	~	[wʳibus]
Target /j/	("re+")	"yellow"	[jɛdou]	~	[wʳijɛdou]
	"	"you"	[ju]	~	[wʳiju]
		"yard"	[jaːb]	~	[wʳijaːb]

phonology. Perhaps, the absence of alternations was associated merely with the structure of English morphology and limitations of the prefix task, rather than the nature of J's phonology.

From these data, four general observations were made about J's phonological system: (a) phonemes used in word-initial position were limited to /m,b,w,j/ (and, of course, vowels); (b) most morphemes were produced without word-initial consonants; (c) a positional constraint limited production of most consonants to postvocalic positions; and (d) an inventory constraint excluded production of [f,v] in all positions. J's use of word-initial consonants was severely restricted and thus served as the primary focus of intervention.

Maximal Opposition Treatment

Experimental Design

In this study, the maximal opposition approach to treatment was implemented within the framework of a single-subject multiple baseline design across 21 sounds. The 21 charted sounds included: /m,b,w,j/, the 4 phonemes J used in word-initial position, /n,p,t,d,k,g,f,v,θ,s,z,ʃ,tʃ,dʒ,h,l/, the 16 phonemes not used in word-initial position, and /r/, which served as a control sound. Production of these 21 sounds was evaluated using a generalization probe measure consisting of a total of 178 words (89 nonprefixed words plus the same 89 words with the prefix "re-" added). Each of the 21 sounds was sampled a minimum of 6 times. Probe items were randomized and elicited in a spontaneous picture-naming task both pre- and post-treatment, as well as at various points throughout treatment.

Multiple baseline designs require stable pretreatment baselines to demonstrate experimental control and effectiveness of treatment. In this study, with 21 sounds charted over time, changes were likely in the pretreatment baselines of at least some of these sounds. Thus, if J's production of an untreated sound changed during baseline, this sound was not selected for subsequent intervention; rather, performance was monitored and the facilitative effects of treatment were examined. From spontaneous baseline changes of this type, it was possible to identify which features or properties of treated sounds J selected as relevant and, further, to examine how J incorporated these relevant dimensions into his sound system. However, of the 21 charted sounds, changes were not expected in the 4 phonemes J already used in word-initial position, namely, /m,b,w,j/. Predictably, these sounds would be produced with 100% accuracy throughout treatment. Similarly, changes were not expected in production of target /r/ since it was characterized by another type of error (i.e., distortions). The phonetic realization of /r/ would likely remain 0% accurate throughout. Changes in the baselines of these sounds would indicate loss of experimental control. Charting 21 sounds, therefore, was consistent with the flexibility of single-subject designs (Connell & Thompson, 1986; Kearns, 1986; McReynolds & Thompson, 1986) and with the recommendation that teaching approaches should be loosely structured (Stokes & Baer, 1977).

Treatment Procedure

Treatment sessions were held twice weekly for 30-min each session. Initially, a pretreatment baseline of all 21 sounds was obtained. One maximal opposition contrast was then selected for treatment based on the multiple distinctions of voice, place, and manner. To illustrate, at the onset of treatment, J only used voiced sounds word-initially; he did not produce a voicing distinction in this position. In terms of place, J primarily used bilabial sounds initially. Also, he only produced the oral-nasal (i.e., /b,w,j/ versus /m/) and stop-glide (i.e., /m,b/ versus /w,j/) manner distinctions. Therefore, it was important that the first maximal opposition be aimed at introducing a voiceless sound produced in a more posterior place of articulation of either the fricative, affricate, or liquid manners. The phoneme /s/ was thus selected for contrast with /m,b,w/, phonemes already used by the child in word-initial position. Other potential treatment candidates considered at this time included /l,ʃ,tʃ/. The phoneme /l/ was not selected for treatment since it is voiced; /ʃ,tʃ/ were not selected since J already used the palatal sound /j/ word-initially.

Actual treatment involved contrasting five picturable word pairs (e.g., "sad"-"mad," "sat"-"mat," "see"-"bee," "suit"-"boot," "sail"-"whale") in first an imitative and then a spontaneous phase of production. During the imitative phase, picture pairs were presented and J was required to name the items following the clinician's verbal model. During the spontaneous phase, the same picture pairs were presented and J named each item without a model. Treatment during the imitative phase was primarily drill; treatment during the spontaneous phase included drill as well as sorting and matching tasks to maintain J's interest and attention. In the sorting task, J spontaneously named picture pairs, placing each picture in its respective sound pile. In the matching task, an array of picture pairs was presented to the child. J selected one picture (e.g., "sad"), named it, and then found its contrasting "match" (e.g., "mad"), naming it as well. Treatment did not involve direct perceptual contrasts of the picture pairs in either the imitative or the spontaneous phase. However, given that the child heard productions of these pairs, incidental perceptual instruction may have been provided. It would be difficult, at best, to determine which specific perceptual cues J attended to during the course of production treatment.

Treatment continued in each phase until J produced word-initial consonants in treatment pairs with 90% accuracy over each of two consecutive 30-min sessions. Upon reaching criterion in both imitative and spontaneous phases, the generalization probe of all 21 sounds was readministered. A second maximal opposition was selected for treatment based on the nature of J's generalization learning. Five new picturable word pairs were then chosen for treatment in both imitative and spontaneous phases of production. Treatment and generalization probes continued in this manner until the child mastered all 16 word-initial sounds. A final generalization probe measure and spontaneous connected speech sample were obtained one week following the completion of treatment.

Reliability

The investigator and two trained listeners (SC, KH) with experience in narrow phonetic transcription served as reliability judges. The investigator and one of the listeners (SC) independently transcribed a portion (20%) of J's pretreatment spontaneous speech sample. Consonant transcriptions were compared point-to-point. Mean transcription reliability was 80% agreement (N = 222 segments). The investigator and the second listener (KH) independently transcribed all of J's responses on repeated administrations of the generalization probe measure. Consonant transcriptions were compared point-to-point. Mean transcription reliability was 96% agreement (N = 1,988 segments; range: 92% to 100% agreement).

Results and Discussion

J received production treatment on three sets of contrasts involving maximal oppositions over the course of three months (23 treatment sessions). Both generalization and overgeneralization data were used to evaluate the effectiveness of a maximal opposition approach to treatment; these data are shown in Table 2 and Figure 1. Table 2 reports percentages of accurate production of the 21 word-initial sounds as sampled on repeated administrations of the probe measure. Figure 1 displays expansions in the range of sounds used in word-initial position, decreases in the range of sounds omitted from word-initial position, and overgeneralizations of certain word-initial consonants. These data were examined with regard to two main questions: What did J learn about specific target sounds and contrasts? And, what did J learn more generally about the class of word-initial consonants?

Insert Table 2 about here

Insert Figure 1 about here

Specific Sounds and Contrasts: Generalization Learning

The first question, knowledge of specific sounds and contrasts, is evaluated with reference to Table 2 and the first column of Figure 1. Recall that, pretreatment, J only used a subset of phonemes word-initially, /m,b,w,j/ (and, of course, vowels). The first maximal opposition that was taught involved contrasting /s/ with /m,b,w/, that is, a voiceless fricative of a more posterior place of articulation versus voiced stops and a glide of a bilabial place of articulation. Following maximal opposition treatment over 8 sessions, J generalized accurate word-initial production to novel words with the treated phoneme, /s/, as well as to other words with untreated phonemes, /n,h/. Generalization to all three phonemes was with 100% accuracy. This pattern of generalization relative to the treated phoneme /s/ suggested that J selected relevant features of contrast along all three dimensions of voice, place, and manner. Specifically, the place features [+coronal] and [+anterior] appeared to be important in generalization to /s,n/ and the voice and manner features [-voice] and [+continuant] in generalization to /s,h/. Interestingly, generalization extended to only those sound classes that J already used in word-initial position, that is, nasals and glides. From this initial generalization learning, we predicted that J would continue to expand the range of phonemes used in word-initial position along any one (or all) of the features [+coronal], [+anterior], [-voice], and [+continuant] by spontaneously learning such consonants as /t,d,z,ʃ,l/.

A second maximal opposition was then selected for treatment. Note that, at this time, J used both word-initial voiced and voiceless sounds, although voiceless sounds were used to a limited degree. Also, he exhibited use of the bilabial, alveolar, palatal, and laryngeal places of articulation. Finally, he relied on four different manners of production: nasals, stops, fricatives,

Table 2

Percentages of Accurate Production of 21 Word-initial Sounds as Sampled on Repeated Administrations of the Generalization Probe Measure

		Pre	Treatment	Sequence		Post
			I	II	III	
Targets			/s/	/tʃ/	/f/	
Contrasts			/m,b,w/	/m,b,s/	/m,b,s,tʃ/	
Word-initial sounds (n=4)	m	100	100	100	100	100
	b	100	100	100	100	100
	w	100	100	100	100	100
	j	100	100	100	100	100
Nonoccurring word-initial sounds (n=16)	n	0	100	100	100	100
	p	0	0	0	100	100
	t	0	0	40	100	100
	d	0	0	100	100	100
	k	0	0	0	0	100
	g	0	0	0	0	100
	f	0	0	0	100	100
	v	0	0	0	25	75
	θ	0	0	0	33	66
	s	0	100	100	100	90
	z	0	0	33	66	100
	ʃ	0	0	20	20	30

Table 2 (cont.)

		Pre	Treatment Sequence			Post
			I	II	III	
Targets			/s/	/tʃ/	/f/	
Contrasts			/m,b,v/	/m,b,s/	/m,b,s,tʃ/	
Nonoccurring word-initial sounds (n=16)	tʃ	0	0	100	100	100
	dʒ	0	0	30	0	80
	h	0	100	100	100	100
	l	0	0	100	100	100
Control sound (n=1)	r	0	0	0	0	0

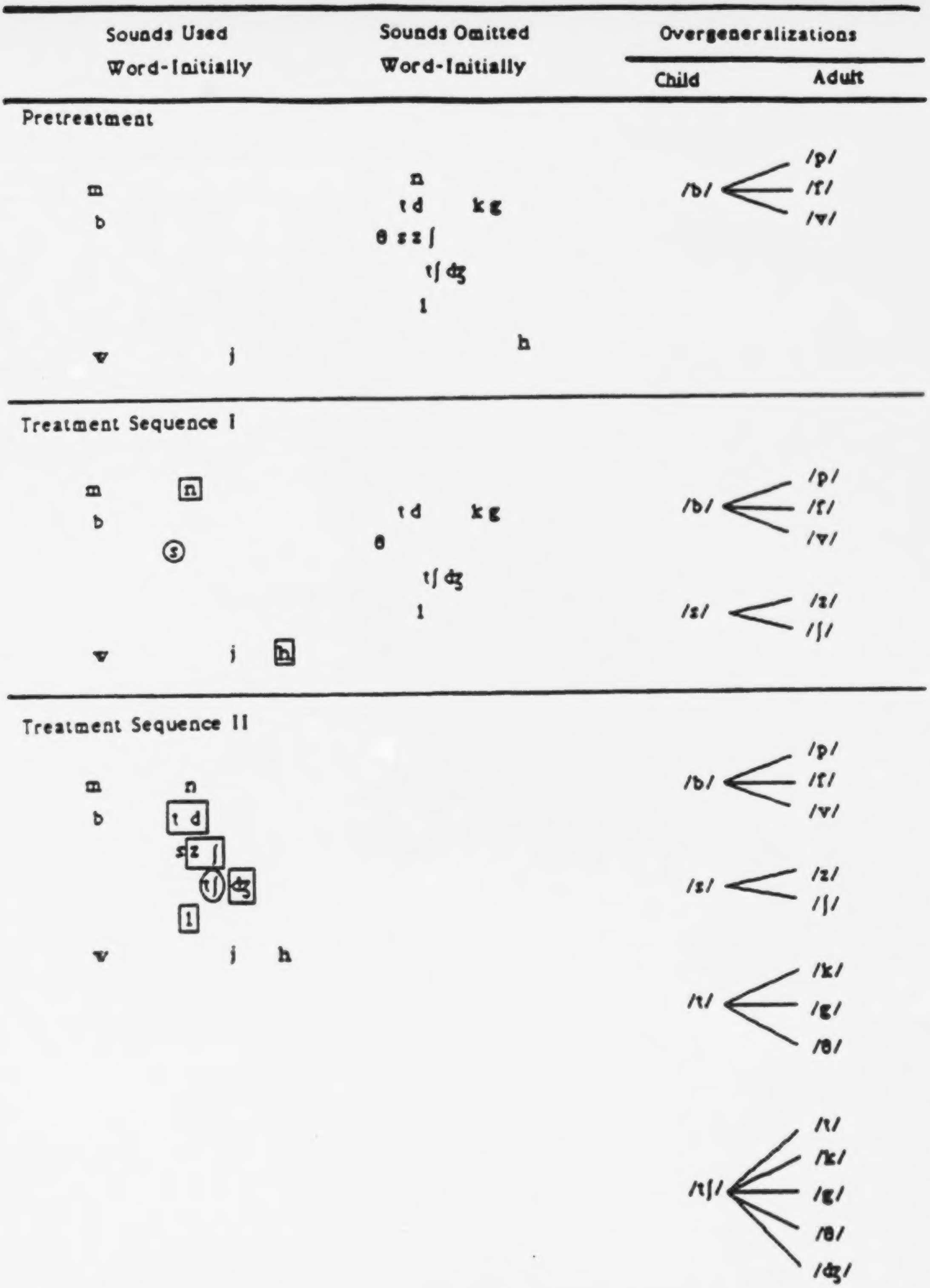
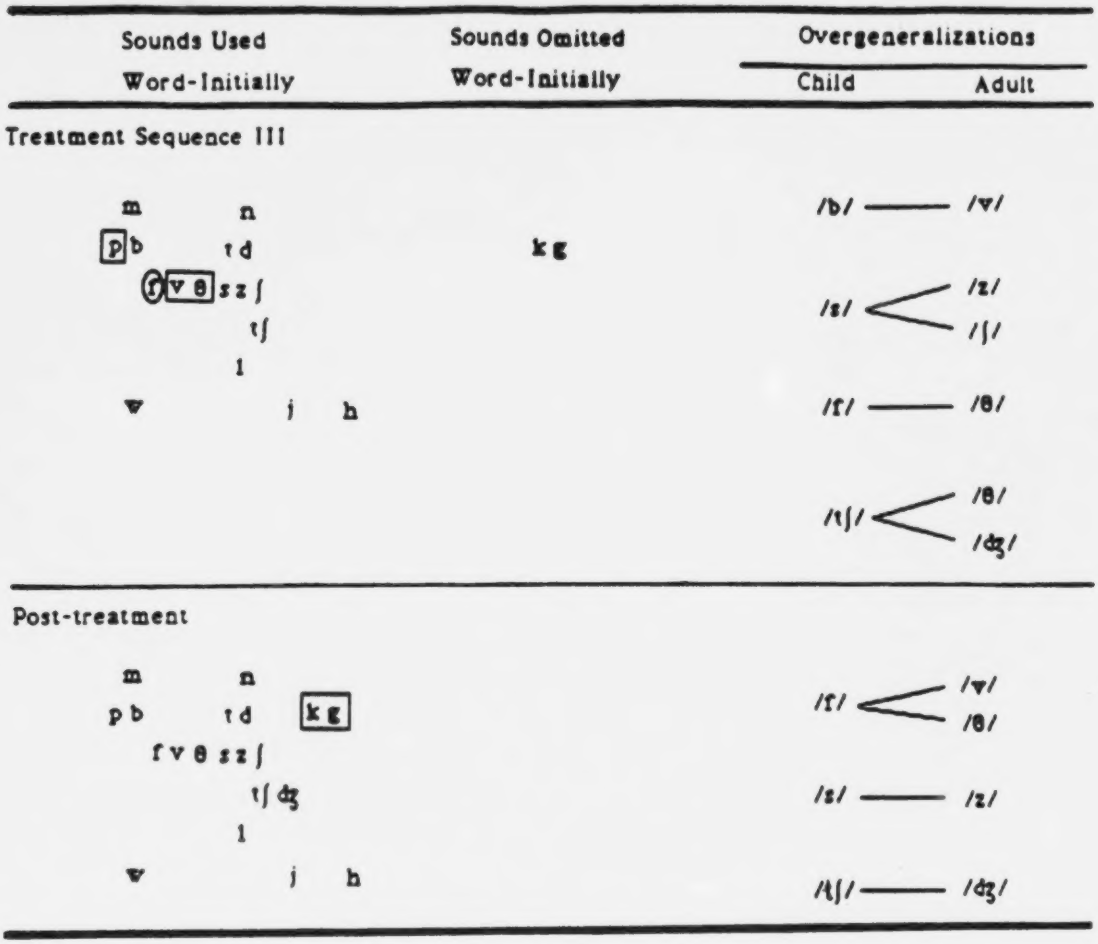

Sounds Used Word-Initially
Sounds Omitted Word-Initially
Overgeneralizations
Child
Adult
Pretreatment
m
b
v j
n
t d k g
θ s z ʃ
tʃ dʒ
l
h
/b/ /p/ /f/ /v/
Treatment Sequence I
m n
b
s
v j h
t d k g
θ
tʃ dʒ
l
/b/ /p/ /f/ /v/
/s/ /z/ /ʃ/
Treatment Sequence II
m n
b t d
s z ʃ
tʃ dʒ
l
v j h
/b/ /p/ /f/ /v/
/s/ /z/ /ʃ/
/t/ /k/ /g/ /θ/
/tʃ/ /t/ /k/ /g/ /θ/ /dʒ/

Figure 1. Phonological restructuring as evidenced by expansions in the range of sounds used in word-initial position, decreases in the range of sounds omitted from word-initial position, and overgeneralizations of certain word-initial consonants. Untreated sounds that were used word-initially are squared; treated sounds are circled.

Figure 1 (cont.)


Sounds Used Word-Initially
Sounds Omitted Word-Initially
Overgeneralizations
Child
Adult
Treatment Sequence III
m n
p b t d
f v θ s z ʃ
tʃ
l
w j h
k g
/b/ — /v/
/s/ < /z/ /ʃ/
/f/ — /θ/
/tʃ/ < /θ/ /dʒ/
Post-treatment
m n
p b t d k g
f v θ s z ʃ
tʃ dʒ
l
w j h
/f/ < /v/ /θ/
/s/ — /z/
/tʃ/ — /dʒ/

and glides. To differentiate further and expand the contrasts in this child's system, it was important to introduce and reinforce voiceless sounds produced in more posterior places of articulation of either the affricate or liquid manners; thus, /tʃ/ was selected for treatment in contrast with /m,b/ and /s/. The reason /s/ was selected for contrast in place of /w/ was to provide the child with additional practice on production of this newly learned sound. Following treatment of this maximal opposition over 5 sessions, J generalized to the word-initial consonants /t,d,z,ʃ,tʃ,dʒ,l/; however, generalization was not complete (100% accurate) in all cases. This generalization pattern relative to the treated phoneme /tʃ/ suggested that J identified a new feature dimension, stridency, as a significant aspect of contrast, as evidenced by his use of word-initial /ʃ,tʃ,dʒ/. Although stridency was first introduced in treatment of /s/, perhaps, treatment of /tʃ/ highlighted further the importance and relevance of this feature for the child. Also, as predicted, J continued to focus on the [+coronal], [+anterior], [-voice], and [+continuant] features. By generalizing to /t,d,z,l/, J likely was incorporating the place features [+coronal] and [+anterior]. The manner feature [+continuant] apparently was expanded through the use of /z,ʃ,l/ word-initially, and the [-voice] feature, through the use of /t,ʃ,tʃ/. Finally, generalization was observed to the existing sound classes of J's system, as well as to new sound classes (i.e., liquids and affricates).

At this point in treatment, J's use of word-initial phonemes was characterized by many finely differentiated oppositions, including a two-way voice contrast (i.e., voiced-voiceless), a five-way manner contrast (i.e., nasal-stop-fricative-affricate-liquid-glide) and a four-way place contrast (i.e., bilabial-alveolar-palatal-laryngeal). In order to complete the full range of word-initial consonants, J needed to differentiate further place of articulation features associated with labiodental (i.e., /f,v/), dental (i.e, /θ/), and velar (i.e., /k,g/) consonants. Thus, the third opposition selected for treatment was /f/ in contrast with /m,b,s/ and /tʃ/. Note that /tʃ/ was also selected for contrast in order to provide J with continued practice on this newly learned sound. Target /f/ was selected over other place distinctions because, potentially, it would strengthen previously treated features that J identified as relevant, namely, [+anterior], [-voice], [+continuant], and [+strident]. While perhaps not intuitive, treatment of /f/ might also elaborate the [-coronal] feature for possible generalization to /k,g/. Following treatment of /f/ over 4 sessions, J generalized accurate word-initial productions to /p,f,v,θ/; generalization was not 100% accurate in all cases. Continued gains were noted, however, in the accuracy of other sounds J previously introduced in word-initial position. From this pattern of generalization relative to the treated phoneme /f/, J apparently attended to only those previously treated relevant dimensions. Although J generalized to /p/, less importance may have been assigned to the [-coronal] feature, as evidenced by his lack of generalization to either /k/ or /g/. At this point, the relatively complete nature of J's word-initial consonant repertoire warranted his dismissal from the maximal opposition treatment program.

A generalization probe administered one week post-treatment indicated that /k,g/ were used in word-initial position with 100% accuracy; perhaps, then, J did identify and elaborate on the [-coronal] feature following treatment of /f/. Also, further improvements were observed in production of other word-initial sounds. Of the 16 charted word-initial sounds, 11 were used with 100% accuracy following treatment of only 3 sets of contrasts. Moreover, four other sounds were used in word-initial position with greater than 65% accuracy. Only one sound, /ʃ/, was used word-initially with less than 50% accuracy following treatment. These observations were also supported in conversational speech. Specifically, the only consistent errors J exhibited in conversational speech post-treatment involved production of targets /f,v/ postvocalically, /r/, and /l/ clusters. Thus, over a relatively short period of intervention involving direct treatment of only three sets of maximally opposed contrasts, J made substantial improvements in the nature of his phonological system. It should also be noted that, throughout treatment, production of the control phoneme /r/ did not improve, nor were changes observed in word-initial production of previously known phonemes, /m,b,w,j/.

These generalization data demonstrated that, for J, a treatment approach based on maximal oppositions was effective in changing and improving the phonological system. The apparent success of this treatment approach may have been associated with J's specific pattern of production involving extensive omissions. It will be important to evaluate further the efficacy of this treatment approach relative to other patterns of production and relative to other methods of contrast treatment.

The generalization data also suggested that J's approach to phonological learning involved building on what was previously learned in treatment. This capitalization on prior learning was evident both in the particular sounds and the extent to which J generalized. That is, relevant features learned by J (e.g., [-voice], [+continuant]) were repeatedly incorporated into sounds generalized later in treatment. Also, gradual improvements in the accurate production of sounds were observed over the course of intervention. J continued to refine his production of word-initial consonants, even after direct treatment of those (or related) sounds.

In addition, J apparently assigned some priority to those oppositions that were treated first. The child seemed to rely on a set or core of features (e.g., [+coronal], [+anterior], [-voice], [+continuant]) in expanding his word-initial repertoire. This observation suggests that, perhaps, first treated oppositions drive or govern the course of later phonological acquisition and learning (cf. Gierut et al., 1987, for a related hypothesis). The role of order in phonological treatment is, of course, subject to experimental test.

Finally, based on patterns of generalization learning, it was possible to generate predictions about those sounds and contrasts that J would spontaneously add to word-initial position and those that would need to be directly taught (cf. Fey & Stalker, 1986). Although target and contrast sounds were selected for direct treatment by the clinician, J's generalization patterns guided this selection process and thus the course of intervention. The maximal opposition approach to phonological treatment seemed to provide J with considerable flexibility and control in choosing the contrasts that would be learned and generalized.

Phonological Restructuring: Overgeneralization Learning

J's more general knowledge of the class of word-initial consonants is evaluated with reference to Figure 1. Figure 1 illustrates the nature and degree of phonological restructuring that occurred in J's phonology over the course of treatment.

A first observation is that the number of consonants omitted from word-initial position decreased substantially following treatment of only one set of maximally opposed contrasts. Although only three new sounds were used accurately in word-initial position at this time, J began marking, albeit incorrectly, the occurrence of many more word-initial consonants. J marked word-initial consonants by overgeneralizing /b/ to target /p,f,v/ and /s/ to target /z,ʃ/. After treatment of the second set of maximal oppositions, all target sounds and morphemes were marked word-initially in some way. At this time, J's error pattern could no longer be characterized by the nonoccurrence of word-initial consonants. The child's original phonological problem of omitting word-initial consonants was no longer of primary concern. Instead, "fine tuning" through continued treatment was needed to bring J's use of word-initial consonants more in line with the target system. This was consistent with the fact that the number and type of oppositions to be learned by the child at this point in remediation (i.e., Treatment Sequence III) was substantially reduced. Treatment contrasts were no longer maximal oppositions but, rather, more finely differentiated minimal distinctions.

A second observation was that the child overgeneralized use of certain phonemes. Overgeneralization data provided supportive evidence and insight into at least two related domains. Specifically, overgeneralizations supported claims of the initial phonological analysis that J did not lexically represent most morphemes for production purposes with word-initial consonants. Overgeneralizations suggested that J treated all omitted sounds as equivalent. When J learned that consonants belong in the initial position, it seemingly did not matter **which** consonant served as the marker. For instance, after learning /tʃ/, J began marking word-initial position (incorrectly) with this sound. He used /tʃ/ to mark word-initial position in target morphemes beginning with such diverse phonemes as /k/ (e.g., [tʃʌp] "cup"), /g/ (e.g., [tʃwəl] "girl") and /θ/ (e.g., [tʃɛʷsi] "thirsty"). If, on the other hand, J lexically represented morphemes for production purposes with target-appropriate word-initial consonants, overgeneralizations of this type would not be expected; instead, generalization would be limited to only those morphemes represented with the same word-initial phoneme. We would expect that, following treatment of /tʃ/, J would mark word-initial position only in other (untreated) /tʃ/ morphemes.

Overgeneralizations also illustrated the nature of restructuring in this child's sound system. These data implied that J formed and changed his phonology in a conceptually-based manner centered on "word-initialness" (as opposed to omissions). While J acquired specific consonants, he also apparently learned about larger units of organization and broader phonological categories. Further evidence of conceptual restructuring based on "word-initialness" comes from data that relate to the acquisition of /f/. Recall that [f,v] were excluded from J's phonetic and phonemic inventories, as accounted for by an inventory constraint. Treatment of /f/ resulted in accurate production and use of this consonant in word-initial position; however, no improvements were observed in production of /f/ in postvocalic positions. This result was particularly interesting in light of the fact that fricatives, in general, and /f/, in particular, are typologically more marked, and presumably more difficult to learn, in word-initial position (Greenberg, Ferguson, & Moravscik, 1978). [2] Given the markedness value of this fricative and the child's overall pattern of postvocalic production, we anticipated generalization of /f/ to all word positions. The lack of postvocalic generalization suggested that J was not so much learning specific consonants as the concept of "word-initialness."

Together, these observations indicated that maximal opposition treatment encouraged J's acquisition of word-initial consonants through conceptualization. J learned and generalized specific consonants; yet, the nature and extent of phonological restructuring suggested that J learned larger phonological and organizational categories. Moreover, these results suggest that quantitative data alone may not represent fully the degree of phonological learning that takes place in a child's sound system during treatment (cf. Elbert & McReynolds, 1979; Leonard & Brown, 1984; Rockman & Elbert, 1984; Weiner, 1981). For J, qualitative changes in the form of marking word-initial position preceded quantitative changes in the form of accurate sound production. For both clinical and research purposes, we must begin to examine and to be sensitive to subtle changes and restructuring and not to underestimate phonological gains by limiting the definition of phonological learning to percentages of accurate sound production.

Approaches to Contrast Treatment

Through this examination of maximal opposition treatment, some of the basic elements and differences of contrast treatment have been highlighted. Whether the focus is on maximal or minimal distinctions, the overall goal of contrast treatment remains the same: to present a conceptual approach to the acquisition of phonemic distinctions in order to reduce the occurrence of homonomy in a child's phonological system. The basic structure, however, of minimal versus maximal contrast treatment is different in at least two ways. One obvious difference, and that which was primarily addressed in this paper, is in degree or breadth of the sound contrasts. In a minimal pair approach, treated distinctions are fine-grained along a focused dimension, as in strident versus nonstrident or voiced versus voiceless. In a maximal opposition approach, distinctions are more global along the broader, multiple dimensions of voice, place, and manner. Within this approach, however, as more oppositions are learned by a child, distinctions become further differentiated eventually leading to minimal contrasts.

A second difference lies in the nature of sounds selected for contrast, as shown in Table 3. In a minimal pair approach, a child is taught to contrast his or her error (i.e., a substituted or omitted sound) with the appropriate target sound. Returning to J's case, a program of minimal pair treatment would have contrasted null (J's error of omission) with relevant target sounds (desired productions) word-initially, as in the potential pairs "at"-"sat," "eat"-"seat," or "ink"-"sink." Within this approach, selection of sounds for treatment is based on a child's phonemic errors relative to the target.

Insert Table 3 about here

In a maximal opposition approach, a child is taught to contrast target sounds that are not used appropriately with those that are currently used in his or her phonological system. For J, the program of maximal opposition compared /s/, a phoneme not occurring in word-initial position, with /m,b,w/, phonemes occurring in this position, as in the pairs "sad"-"mad," "see"-"bee," "sail"-"whale." Within this approach, selection of treatment sounds is based on occurrences and nonoccurrences in the child's phonological system relative to the target.

One other form of contrast treatment comes to mind in light of this discussion, which will be called treatment of the empty set. In this approach, a child is taught to contrast two sounds that do not occur in his or her phonemic inventory. Returning to J, a program of empty set treatment

Table 3

Sounds Selected for Contrast Using Different Approaches to Contrast Treatment

	Types of Contrast Treatment		
	Minimal Pair	Maximal Opposition	Empty Set
Sounds of the Target Phonology	X	X	X X
Sounds of the Child's Phonology			
Errored Aspects of the System	X		
Correct Aspects of the System		X	

would have contrasted, for example, /s/ with /tʃ/ in the potential pairs "sip"-"chip," "sick"-"chick," or "Sue"-"chew," because neither phoneme was used in word-initial position. In this approach, sounds selected for contrast focus only on what a child has yet to learn about the target phonology.

One important research question that remains to be asked is whether such variations in the format of contrast treatment will result in empirical differences among the treatment approaches. For example, different approaches to contrast treatment may influence the nature, extent, and type of generalization that a child will display following treatment. Also, certain forms of contrast treatment may be more appropriate for some children than others. The nature of the error pattern may potentially contribute to the effectiveness of the various treatment approaches. A child who displays a relatively complete phonemic inventory with only one or two errors may be more appropriately suited for a program of minimal pair distinctions; whereas, a child like J who displays extensive gaps in the system may benefit from a program of maximal oppositions or treatment of the empty set. Children with inconsistent errors or variable productions may also be well-suited for a maximal opposition approach given that several feature dimensions are contrasted simultaneously and that the child may focus attention on more than one distinction. Finally, individual learning styles may limit treatment effectiveness. For some children, narrower, minimal distinctions may be more difficult to attend to and to master than broader, multiple distinctions; such has been the case with some specifically language-impaired children learning syntax (Connell, 1986). Only continued comparative study of contrast treatment approaches will provide us with answers to such important basic and applied research questions.

Footnotes

1 The term "feature," as used herein, is consistent with the Chomsky-Halle distinctive feature framework (Chomsky & Halle, 1968), although other cross-classificatory feature systems (e.g, Jakobson, Fant, & Halle, 1951; Ladefoged, 1975) would be equally applicable. Also, the term "phoneme" refers generally to those distinctive properties of sounds that are used to signal meaning differences in a language. When "phoneme" is used in reference to production, it should be interpreted as production of a sound associated with a particular phoneme.

2 Typological markedness is a linguistic phenomenon that identifies a relationship among sounds, such that the occurrence of one sound in a language predicts the occurrence of other sounds in that same language. The predicting or implying sound is "marked" relative to the predicted or implied "unmarked" sound. For example, if a language has voiced obstruents, it will also have voiceless obstruents; voiced obstruents are marked relative to voiceless obstruents.

References

Bates, E. (1976). *Language in context*. New York: Academic Press.

Bates, E., & MacWhinney, B. (1982). Functionalist approach to grammar. In E. Wanner & Gleitman, L.R. (Eds.), *Language acquisition: The state of the art* (pp. 173-218). Cambridge, England: Cambridge University Press.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Connell, P.J. (1986). Teaching subjecthood to language-disordered children. *Journal of Speech and Hearing Research*, *29*, 481-492.

Connell, P.J., & Thompson, C.K. (1986). Flexibility of single-subject experimental designs. Part III: Using flexibility to design or modify experiments. *Journal of Speech and Hearing Disorders*, *51*, 214-225.

Costello, J., & Onstine, J.M. (1976). The modification of multiple articulation errors based on distinctive feature theory. *Journal of Speech and Hearing Disorders*, *41*, 199-215.

Crocker, J. (1969). A phonological model of children's articulation competence. *Journal of Speech and Hearing Disorders*, *34*, 203-213.

Dinnsen, D.A. (1984). Methods and empirical issues in analyzing functional misarticulations. In M. Elbert, D.A. Dinnsen, & G. Weismer (Eds.), *Phonological theory and the misarticulating child* (ASHA Monograph 22, pp. 5-17). Rockville, MD: ASHA.

Elbert, M. (1984). The relationship between normal phonological acquisition and clinical intervention. In N.J. Lass (Ed.), *Speech and language: Advances in basic research and practice: Vol. 10*. (pp. 111-139). New York: Academic Press.

Elbert, M., Dinnsen, D.A., & Powell, T.W. (1984). On the prediction of phonologic generalization learning patterns. *Journal of Speech and Hearing Disorders*, *49*, 309-317.

Elbert, M., & Gierut, J.A. (1986). *Handbook of clinical phonology: Approaches to assessment and treatment*. San Diego: College-Hill Press.

Elbert, M., & McReynolds, L.V. (1979). Aspects of phonological acquisition during articulation training. *Journal of Speech and Hearing Disorders*, *44*, 459-471.

Elbert, M., Rockman, B., & Saltzman, D. (1980). *Contrasts: The use of minimal pairs in articulation training*. Austin, TX: Exceptional Resources.

Ferguson, C.A., & Farwell, C.B. (1975). Words and sounds in early language acquisition. *Language*, *51*, 419-439.

Ferguson, C.A., & Macken, M.A. (1980). Phonological development in children: Play and cognition. *Papers and Reports on Child Language Development*, *18*, 138-177.

Ferguson, C.A., Peizer, D.B., & Weeks, T.E. (1973). Model-and-replica phonological grammar of a child's first words. *Lingua*, *31*, 35-65.

Ferrier, E., & Davis, M. (1973). A lexical approach to the remediation of final sound omissions. *Journal of Speech and Hearing Disorders*, *38*, 126-130.

Fey, M.E., & Gandour, J. (1982). Rule discovery in phonological acquisition. *Journal of Child Language*, *9*, 71-81.

Fey, M.E., & Stalker, C.H. (1986). A hypothesis-testing approach to treatment of a child with an idiosyncratic (morpho)phonologial system. *Journal of Speech and Hearing Disorders*, *51*, 324-336.

Gierut, J.A. (1985). *On the relationship between phonological knowledge and generalization learning in misarticulating children*. Doctoral dissertation, Indiana University, Bloomington. (Distributed by the Indiana University Linguistics Club, 1986)

Gierut, J.A. (1986). On the assessment of productive phonological knowledge. *NSSLHA Journal*, *14*, 83-101.

Gierut, J.A., Elbert, M., & Dinnsen, D.A. (1987). A functional analysis of phonological knowledge and generalization learning in misarticulating children. *Journal of Speech and Hearing Research*, *30*, 462-479.

Goldman, R., & Fristoe, M. (1969). *Goldman-Fristoe test of articulation*. Circle Pines, MN: American Guidance.

Greenberg, J.H., Ferguson, C.A., & Moravscik, E.A. (Eds.). (1978). *Universals of human language: Vol. 2. Phonology*. Stanford, CA: Stanford University Press.

Ingram, D. (1976). *Phonological disability in children*. New York: Elsevier.

Jakobson, R. (1968). *Child language, aphasia, and phonological universals*. (A.R. Keiler, Trans.). The Hague: Mouton. (Original work published 1941)

Jakobson, R., Fant, G., & Halle, M. (1951). *Preliminaries to speech analysis*. Cambridge, MA: MIT Press.

Johnston, J.R. (1982). The language disordered child. In N. Lass, L. McReynolds, J. Northern, & D. Yoder (Eds.), *Speech, language, and hearing: Vol. 2*. (pp. 780-801). Philadelphia: W.B. Saunders.

Kearns, K.P. (1986). Flexibility of single-subject experimental designs. Part II: Design selection and arrangement of experimental phases. *Journal of Speech and Hearing Disorders*, *51*, 204-214.

Ladefoged, P. (1975). *A course in phonetics*. New York: Harcourt Brace Jovanovich.

Leonard, L.B. (1981). Facilitating linguistic skills in children with specific language impairment. *Applied Psycholinguistics*, *2*, 89-118.

Leonard, L.B., & Brown, B.L. (1984). Nature and boundaries of phonologic categories: A case study of an unusual phonologic pattern in a language-impaired child. *Journal of Speech and Hearing Disorders*, *49*, 419-428.

Leopold, W.F. (1947). *Speech development of a bilingual child: A linguist's record: Vol. 2. Sound-learning in the first two years*. Evanston, IL: Northwestern University.

Macken, M.A., & Ferguson, C.A. (1983). Cognitive aspects of phonological development: Model, evidence and issues. In K.E. Nelson (Ed.), *Children's language: Vol. 4*. (pp. 255-282). Hillsdale, NJ: Erlbaum.

McReynolds, L.V., & Bennett, S. (1972). Distinctive feature generalization in articulation training. *Journal of Speech and Hearing Disorders*, *37*, 462-470.

McReynolds, L.V., & Thompson, C.K. (1986). Flexibility of single-subject experimental designs. Part I: Review of the basics of single-subject designs. *Journal of Speech and Hearing Disorders*, *51*, 194-203.

Menn, L. (1976). *Pattern, control, and contrast in beginning speech: A case study in the development of word form and word function*. Doctoral dissertation, University of Illinois, Champaign-Urbana. (Distributed by the Indiana University Linguistics Club, 1978)

Philips, B.J., & Harrison, R.J. (1969). Articulation patterns of preschool cleft palate children. *Cleft Palate Journal*, *6*, 245-253.

Priestly, T.M.S. (1977). One idiosyncratic strategy in the acquisition of phonology. *Journal of Child Language*, *4*, 45-66.

Rockman, B.K., Dinnsen, D.A., & Rowland, E. (1983, November). *A case of initial consonant deletion: Phonological issues and implications*. Paper presented at the American Speech-Language-Hearing Association Convention, Cincinnati, OH.

Rockman, B.K., & Elbert, M. (1984). Untrained acquisition of /s/ in a phonologically disordered child. *Journal of Speech and Hearing Disorders*, *49*, 246-254.

Schwartz, R.G., & Leonard, L.B. (1982). Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language*, *9*, 319-336.

Stokes, T.F., & Baer, D.M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis*, *10*, 349-367.

Velten, H.V. (1943). The growth of phonemic and lexical patterns in infant language. Language, 19, 281-292.

Vihman, M. (1981). Phonology and the development of the lexicon. Journal of Child Language, 8, 239-265.

Weber, J.L. (1970). Patterning of deviant articulation behavior. Journal of Speech and Hearing Disorders, 35, 135-141.

Weiner, F. (1981). Treatment of phonological disability using the method of meaningful minimal contrast: Two case studies. Journal of Speech and Hearing Disorders, 46, 97-103.

Winitz, H. (1975). From syllable to conversation. Baltimore: University Park Press.

Zimmerman, I.L., Steiner, V.G., & Pond, R.E. (1979). Preschool language scale-Revised. Columbus, OH: Charles Merrill.

Appendix

Probe Items

Items were elicited both as nonprefixed and prefixed (i.e., "re-") forms to evaluate J's production of word-initial sounds.

/m/	/n/
mud	knife
mouth	nose
mother	nail
mouse	
moon	

/p/	/b/	/t/	/d/	/k/	/g/
pig	big	tear	duck	cup	gum
pie	book	tub	deer	cut	girl
pants	bed	toes	door	coat	gun
peach	bus	tail	dog	comb	goat
paint	boot	tooth			

/f/	/v/	/s/	/z/	/θ/	/ʃ/
fat	van	soup	zebra	thumb	shave
face	vase	soap	zipper	thief	shoe
fire	vanilla	sock	zoo	thirsty	shirt
fish	vacuum	santa			shovel
five		sun			shampoo

/tʃ/	/dʒ/	/w/	/j/	/h/
chair	jelly	watch	yellow	hide
cheese	jump	window	you	hug
chip	jeep	wash	yard	hill
	jail	wave		hat
	juice			house

/r/	/l/
read	laugh
rain	leaf
run	light
ride	ladder
	leg

[RESEARCH ON SPEECH PERCEPTION Progress Report No. 13 (1987) Indiana University]

The Effects of Semantic Context on Voicing Neutralization*

Jan Charles-Luce

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, IN 47405

---

*Work supported by NIH Training Grant No. NS-07134-09 and NIH Research Grant No. NS-12179-11. I would like to thank Daniel Dinnsen, David Pisoni, Robert Port, and Linda Schwartz for their comments on an earlier version of this paper and Paul Luce for his comments and for his help in writing the stimulus presentation and data management programs. I also thank Maria Rosa Lloret Romanach and Thomas Walsh for their help in constructing the stimulus materials. Portions of this study were presented at the annual Linguistics Society of America meeting, Dec. 1986.

Abstract

The present study examined regressive voice assimilation in Catalan in an attempt to determine a systematic explanation of complete versus incomplete voicing neutralization. Two types of contexts were constructed. In one type, semantic information was present to bias the target words. In the other type, no semantic information was present to bias the target words. The results show that neutralization is complete in the semantically biasing context, but it is incomplete in the non-semantically biasing context. These findings suggest that phonological processes do not operate in an autonomous module but rather are part of an interactive linguistic system.

The Effects of Semantic Context on Voicing Neutralization

The present study of Catalan is aimed at determining the effects of semantic context on the putative neutralization of an underlying voice contrast. Earlier studies have provided empirical evidence that word-final devoicing does not always result in complete acoustic neutralization of the underlying voice contrast. That is, underlying voiced and voiceless stops are phonetically realized as voiceless in word-final position, resulting in the loss of the contrast. These previous studies examining the neutralization of the voice contrast in Catalan, German, and Polish have been fairly neutral with respect to various aspects of linguistic information that might bias the test words. Typically, minimal pairs have been produced in isolation [Port and O'Dell, 1985] or in single sentence frames [Dinnsen and Charles-Luce, 1984; Charles-Luce, 1985; Slowiaczek and Dinnsen, 1985]. Minimal pairs embedded in sentence frames were essentially unconstrained in their syntactic and semantic occurrence. Thus, these previous studies have not considered how other levels of linguistic information may affect neutralization.

The present study demonstrates that neutralization rules are not abstract processes applying singularly or in conjunction with other phonological processes in an autonomous phonological module. Rather neutralization processes are part of an interactive linguistic system, in particular affected by the degree of semantically biasing information preceding the putative neutralization event.

Contrary to the majority of studies showing incomplete voicing neutralization, Fourakis and Iverson [1984] have reported that neutralization is complete in German. They found no differences in vowel duration preceding final stops or in stop closure duration that distinguished underlying voicing. They employed a verbal conjugation paradigm in which subjects produced the principal parts of German strong verbs. The target words were always the uninflected second principal part. In this form, stops occur word-finally and, therefore, word-final devoicing putatively should neutralize the underlying voice contrast. (Properly speaking, the domain of the devoicing rule is syllable-final in German [Moulton, 1962].) For example, subjects were given the infinitival form *meiden* and asked to produce the three principal forms: *meiden* "to avoid", *mied* "avoided", and *mieden* "have avoided." *Mied* is the target word, putatively realized as [mi:t]. Fourakis and Iverson claim that they found no differences in underlying voicing because "the focus of pronunciation is disguised" [p. 149] and, moreover, that their task provided a more natural situation for testing word-final neutralization.

It should be noted, however, that the test words were not minimal pairs. For example, in the near minimal pair *riet* and *mied*, the initial segments are not identical. Because of coarticulatory effects, different initial segments may adjust the timing of the following vowel (as well as the actual articulatory gestures) that is unrelated to vowel duration as a correlate to voicing [cf. House and Fairbanks, 1953; Lindblom, 1983]. Nonetheless, their claim about the role of a natural situation in determining the extent of neutralization should not go unheeded. However, Fourakis and Iverson's recognition of the importance of a natural situation does not provide an explanation for their "incomplete" neutralization results. The explanation lies in the specific elicitation task they employed. During the conjugation task, subjects had accessed the underlying morpheme and were, therefore, processing all the linguistic information associated with that particular morpheme. Because the verbal paradigm was unique to the word, the linguistic situation afforded neutralization. There was no ambiguity in the mind of the speakers as to the word required to fulfill a particular conjugation.

The related question for the present investigation is what effect does the presence or absence of semantically biasing information have on the neutralization of the voice contrast. By setting up contexts that simulate more natural conversational situations, one can examine this question. It has been shown that comprehension of text is impaired when a semantic context has not been established previously [Bransford and Franks, 1971; Dooling and Lachman, 1971; Bransford and Johnson, 1972]. Futhermore, the production of words ultimately reflects the decisions of a speaker at the semantic level [Lieberman, 1963]. Production studies have shown that words and segments are reduced and less precisely articulated in syntactic and semantically correct contexts, but they are less reduced and more precisely articulated in anomalous and ungrammatical contexts [Lieberman, 1963]. Similarly, Charles-Luce and Walker [1981] found that the duration of words are longest when they were read in ungrammatical sentences, shortest in grammatical sentences, and intermediate in anomalous sentences [cf. Miller and Isard, 1963].

Perceptually, the intelligibility of excised words decreases as a function of the redundancy of semantic information [Pollack and Pickett, 1963; 1964]. For example, Lieberman [1963] had subjects listen to the word *lender* excised from a redundant context ("Neither a borrower nor a lender be.") and from a non-redundant context ("Never listen to a man who wants to be a lender."). In the redundant context, *borrower* sets up a semantic expectation for *lender*. This semantically biasing information is absent in the non-redundant context. Percent identification was higher for words excised from the non-redundant context relative to the redundant context. Lieberman concludes that speakers produce words with less care when they know that listeners will use the context to identify the words [see, also, Miller, Heise, and Lichten, 1950].

Thus, there is evidence suggesting that articulation of words is affected differentially by the presence and absence of higher levels of linguistic information and that the degree of preciseness of articulation is inversely proportional to the presence of semantic information [Lieberman, 1963]. The effect of semantic information is to reduce the acoustic information necessary for identifying the word and accessing its meaning from the lexicon. In natural situations where semantic context is available to indicate the intended lexical item, and, therefore, meaning, the individual acoustic events may be less important as cues to the listener. However, when this top-down semantic information is lacking, then the bottom-up acoustic events may be sufficient cues for communicating the intended word to the listener.

The purpose of this investigation is to examine how the presence and absence of semantically biasing information affects the phonological neutralization processes involving the voicing of word-final stops. The particular language of investigation is Catalan and, in particular, the dialect spoken around Barcelona, Spain. An attractive feature of Catalan is the lack of a word-final orthographic distinction that corresponds with an underlying voicing distinction. Thus, no argument can be made that incomplete neutralization results from speakers' hypersensitivity to a grapheme/phoneme correspondence [cf. Fourakis and Iverson, 1984].

Minimal pairs of CVC words were produced in two assimilatory environments (voiced and voiceless). In assimilatory environments, word-final Catalan stops putatively assume the voicing of the following consonant [Mascaro, 1978; Wheeler, 1979; DeCesaris, 1980]. Thus, both underlying voiced and voiceless stops become voiced in a following word-initial voiced consonant environment or voiceless in a following word-initial voiceless consonant environment. It is hypothesized that words will be less precisely articulated when semantically biasing information is present. Consequently, differences in underlying voicing may not be found because, in this context, the acoustic events are secondary events in signaling the intended meaning. Thus, differences in morphemic representations may not be revealed because the underlying voice contrast can afford to be neutralized. Furthermore, it is hypothesized that words will be more precisely articulated when semantically biasing information is absent. Therefore, differences in underlying voicing may be found because the acoustic events are necessary cues in the absence of semantically biasing information. In this case, differences in morphemic representations may be revealed because the acoustic obliteration of the underlying voice contrast may be disadvantageous to speaker/hearer communication because some ambiguity may result.

Method

Five minimal pairs of words were selected as stimuli. The criteria for selecting the five minimal pairs were based solely on the ability to semantically constrain both members of the minimal pairs in the most efficacious manner, as described below. Table I presents the five minimal pairs.

Insert Table I about here

Each test word occurred in two types of contexts. Examples of these contexts are shown in Table II and will, henceforth, be referred to as: (1) the semantically biasing context and (2) the non-semantically biasing context.

Insert Table II about here

In the first context (Paragraphs 1a and 1b), a test word was embedded in the last sentence of a two-sentence paragraph that syntactically *and* semantically constrained the lexical category and the meaning of the test word. Importantly, in this context, the words *duquessa* "duchess", *marit* "husband", and *ducat* "dukedom" semantically biased the test word *duc* "duke". In the second context (Paragraphs 2a and 2b), the test word again was embedded in the last sentence of a two-sentence paragraph, but the last sentence *only* syntactically constrained the lexical category of the test word. No preceding lexical items semantically biased the choice of the test word *duc* "duke". Thus, test words were syntactically constrained in both contexts, but only semantically constrained in the first context.

Furthermore, the test words were always preceded by 13 syllables in the last sentence. This manipulation was made to minimize any possible effects of differences in overall word duration in the test words. Such differences

Table I

The five minimal pairs used in this study. The phonetic and underlying representations are given for each word, as well as English glosses.

	Phonetic Representation	Underlying Representation	English Gloss
1.	[rrik]	/rrik/	'rich' masc.
	[rrik]	/rrig/	'I laugh' Pres. Ind.
2.	[duk]	/duk/	'duke'
	[duk]	/dug/	'I carry' Pres. Ind.
3.	[fat]	/fat/	'fate'
	[fat]	/fad/	'tasteless' masc.
4.	[sɛk]	/sɛk/	'dry' masc.
	[sɛk]	/sɛg/	'I sit down' Pres. Ind.
5.	[sɛt]	/sɛt/	'seven'
	[sɛt]	/sɛd/	'thirst'

Table II

Example of the two types of semantic contexts for the test word *duc* "duke". All test words occurred in two types of context and, within each context type, they occurred in two environments: (a) voiceless assimilatory and (b) voiced assimilatory.

1. *Semantically Biasing Context*:

La duquessa i el seu marit viuen a un gran ducat. La

duquessa vella esta ben casada amb el ____ (1a) duc [-voice].

(1b) duc [+voice].

"The duchess and her husband live in a large dukedom. The old duchess is happily married to the duke."

2. *Non-semantically Biasing Context*:

Sempre ens ho passem be anant al parc. Ahir va ploure

molt fort i varem veure el ____ (2a) duc [-voice].

(3b) duc [+voice].

"We always enjoy going to the park. Yesterday it rained very hard and we saw the duke."

might arise from an early versus late occurrence in the last sentence of a paragraph.

In addition to the type of context, the test words occurred in two assimilatory environments within each of the two context types. [1] In Paragraphs 1a and 2a, the test words occurred in an environment with a following word-initial apico-alveolar voiceless fricative [s]. In Paragraphs 1b and 2b, the test words occurred in an environment with a following word-initial apico-alveolar voiced multiple-trill [rr]. These specific assimilatory consonants were selected because they provide two of the few phonetic segments that do not also trigger other types of assimilation in the word-final stops of these Catalan test words. Although the following initial segments are not identical in manner, they are identical in place. Most important for this study, they differ in voicing. If neutralization is complete, then underlying voiced and voiceless word-final stops should be phonetically realized as voiceless preceding the word-initial [s] but as voiced preceding the intitial [rr].

To recapitulate, each test word occurred in each of the two context types: (1) semantically biasing and (2) non-semantically biasing. Within a context type, each test word occurred in each of the two assimilatory environments: (1) following voiceless consonant (henceforth, voiceless assimilatory) and (2) following voiced consonant (henceforth voiced assimilatory).

Five repetitions of each test word in each type of semantic context and in each environment were read by each of the five subjects. This resulted in 200 experimental items [5 minimal pairs x 2 underlying representations x 2 paragraph types x 2 environments x 5 repetitions] for each subject. [2] In addition, 200 filler paragraphs were presented for subjects to read. For the filler paragraphs only, subjects were presented with true/false questions to answer about some word or idea in a filler paragraph that had just been presented to them. Examples of these true/false questions are presented in Table III.

Insert Table III about here

This procedure was intended to distract the subjects from the focus of the experiment and to force them to read all experimental and non-experimental paragraphs for comprehension [cf. Aaronson and Scarborough, 1976]. Subjects did not know on which of the paragraphs they would be asked questions.

All experimental and non-experimental items were fully randomized by computer and each subject received a different randomization of the total 400 paragraphs. All instructions were presented to subjects in Catalan. This was to ensure that all subjects, especially those just learning English, understood the task. They were also intended to help subjects re-acclimate themselves to their native language in the immediately surrounding English-speaking environment.

Each paragraph was presented one at a time on a CRT monitor, positioned at eye level in front of the subject. The words APUNT PER A COMENCAR 'ready to begin' occurred in the center of the CRT screen. When the subject was

Table III

Example of a filler paragraph and a corresponding true/false question.

Filler Paragraph

M'encanta ballar. Ho se ballar tot menys un vals noble.

"I like to dance. I know how to dance every kind of dance except the noble waltz."

True/False Question

El vals es un ball noble.   (vertader/fals?)

"The waltz is a noble dance."

ready to begin reading, s/he pressed a button labeled PER A CONTINUAR 'continue' on a response box in front of her/him and below the CRT monitor. A two-sentence paragraph would then appear in the center of the CRT screen. After the subject had read the paragraph aloud, s/he would again press the button labeled PER A CONTINUAR 'continue'. If the paragraph was an experimental paragraph, then the next paragraph would immediately appear in the center of the screen for the subject to read aloud. If the paragraph was a filler paragraph, then the word PREGUNTA 'question' would flash in the center of the screen. This signaled the subject that a true/false question about the paragraph they had just read aloud was about to appear. The question would then appear in the center of the screen. After the subject had decided whether the correct answer was true or false, s/he pressed the corresponding button, labeled VERTADER 'true' or FALS 'false' on the response box. Subjects were instructed to continue in this manner until the word DESCANS 'rest' appeared on the CRT screen.

Before the experimental recording session begin, subjects were presented with eight practice paragraphs, four of which they had to answer true/false questions and four of which they did not. The experimental recording session began after giving the subjects an opportunity to ask questions.

Five blocks of paragraphs were presented during the experimental recording session, allowing the subjects to have four breaks and allowing for the experimenter to change audio tapes. Blocks one and five had 90 paragraphs each and blocks two through four had 140 paragraphs each. The recording session ended when the words LA FI 'the end' appeared in the center of the CRT screen at the end of the fifth block.

All utterances were recorded in a sound attenuated booth (IAC model 401A) using an Electro-Voice D054 microphone and an Ampex AG-500 tape recorder. In addition, a high-pass filter was employed during recording to filter out extraneous room noise at 60 Hz and below. Sentences and prompts were presented on a CRT monitor (GBC MV-10A). The CRT monitor and response box were interfaced to a PDP 11/34 computer for presentation of the stimuli.

Subjects

Five adult native speakers (four female and one male) of Standard Catalan (the Eastern dialect) served as paid subjects. Three speakers were born in Barcelona, Spain and were still permanent residents of Barcelona. Two speakers were born in Girona, Spain, a town north of Barcelona but still within the linguistic bounds of the Eastern Catalan dialect. At the time of testing, these speakers were also permanent residents of Barcelona. Although today all speakers of Catalan are also speakers of Castilan Spanish, Catalan was the language spoken in the subjects' home and was the first language spoken by all five subjects. No subject reported a history of speech or hearing disorders.

Measurements

Test utterances from each of the five subjects were low-pass filtered at 4.8 kHz and digitized at a sampling rate of 10 K samples per second, via a 12-bit analog-to-digital converter. Measurements were made from a visual waveform display using a digital waveform editor [see Luce and Carrell, 1981].

For each test word, three measurements were made: (1) vowel duration preceding the word-final stop, (2) voicing during closure of the final stop, and (3) closure duration of the final stop. The segmentation criteria for these measurements were as follows:

(1) Vowel duration. For all test words, except the minimal pair /rrik/-/rrig/, vowel duration was defined as the interval from onset of periodicity in the waveform to a marked decrease in amplitude in the waveform and/or change in the shape of the periodic waveform. For the test words /rrik/ and /rrig/, consistently segmenting the word-initial voiced apico-alveolar trill from the following vowel proved problematic because of the variation among subjects' productions of the trill. In particular, it was difficult to establish a consistent criterion that distinguished between the offset of the last vibrating movement of the trill and the onset of the vowel. Thus, for this pair, vowel duration included the word-initial [rr] and following vowel. The onset of the initial [rr] was determined at the juncture between a decrease in amplitude in the smooth periodic waveform of the preceding nasal or vowel and an increase in amplitude for the first vibration of the [rr], as well as a characteristically more complex waveform corresponding to each vibrating movement of the trill.

(2) Voicing during closure duration. Voicing during closure of the final stops was defined as the interval representing glottal pulsing in the closure constriction of the stop, as indicated by a low amplitude periodic waveform. It was measured from the offset of the vowel duration (see above) until energy was no longer detected in the waveform. In cases in which voicing during closure lasted throughout the entire closure of the final stop, the duration of voicing during closure was identical to closure duration (see below).

(3) Stop closure duration. Closure duration for the final stops was defined as the interval from a marked decrease in amplitude of the preceding vowel to onset of the release burst of the final stop, as indicated by a high energy spike in the waveform. In the case of velar stops that sometimes had a double release (one spike of energy followed almost immediately by a second spike of energy), closure duration was measured from offset of the vowel to onset of the first spike of energy.

If differences in underlying voicing are found, the expected direction of the durational differences associated with voiced and voiceless stops should be as follows [cf. Chen, 1970; Ohala, 1983, and references therein]:

(1) Vowels are longer preceding voiced stops relative to voiceless stops,
(2) voicing during closure (length of glottal pulsing) is longer for voiced stops relative to voiceless stops and/or,
(3) closure duration is longer for voiceless stops relative to voiced stops.

Furthermore, it is hypothesized that if underlying voicing is neutralized and regressive voice assimilation applies, then, as correlates of voicing, the three temporal intervals measured should exhibit the durational patterning as predicted below. Obviously, not all the temporal intervals may show voice assimilation, but one or more of the following three intervals may reflect at least partial assimilation in the expected directions.

(1) Vowel duration averaged across underlying voiced and voiceless stops is longer in a voiced assimilatory environment relative to a voiceless assimilatory environment,
(2) voicing during closure averaged across underlying voicing will be longer in a voiced assimilatory environment relative to a voiceless assimilatory environment and/or,
(3) closure duration averaged across underlying voicing will be longer in a voiceless assimilatory environment relative to a voiced assimilatory environment.

Results

For each of the three temporal measurements, repetitions for each test word were averaged within each subject. For each type of context, three-way [underlying voicing x environment x minimal pair] repeated measures analyses of variance were performed separately on mean vowel duration, voicing during closure, and closure duration.

Because this study is concerned with how the presence and absence of semantically biasing information affects the neutralization of the final voice contrast in three environments, only main effects of underlying voicing and environment will be discussed. In addition, only significant interactions involving underlying voicing will be discussed.

Vowel Duration

Table IV shows the mean durations collapsed across lexical items for vowel duration preceding underlying voiced and voiceless word-final stops. The results for the semantically biasing context and the non-semantically biasing context are presented in the top and bottom panels, respectively. Left to right, the columns present the results for the voiceless and voiced assimilatory environments. All durations are in milliseconds.

Insert Table IV about here

For the semantically biasing context (top panel), no significant main effect of underlying voicing was found for vowel duration [$F(1,4) = 7.05$; $p < 0.06$]. However, a significant two-way interaction of underlying voicing and minimal pair [$F(1,4) = 3.33$; $p < 0.04$] and a significant three-way interaction of underlying voicing, environment, and minimal pair [$F(8,32) = 2.80$; $p < 0.02$] were obtained.

One-way analyses of variance performed on the three-way interaction revealed that underlying voicing was distinguished in only two minimal pairs, each in a different assimilatory environment. Figure 1 shows the results for each minimal pair. Mean vowel duration is presented as a function of the minimal pairs. The open bars represent the results for the underlying voiced stops and the filled bars the results for the underlying voiceless stops. The top and bottom panels represent the results for the minimal pairs produced in

Table IV

Mean vowel durations (ms) preceding underlying voiceless and voiced stops produced in the semantically biasing context (top) and in the non-semantically biasing context (bottom) and in each of the two environments: (1) following voiceless assimilatory environment (/_C#[-voice]) and (2) following voiced assimilatory environment (/_C#[+voice]).

Vowel Duration

	Semantically Biasing Context	
	/___C#[-voice]	/___C#[+voice]
UR [-voice]	89	96
UR [+voice]	94	104
Mean	(92)	(100)
	Non-Semantically Biasing Context	
UR [-voice]	85	91
UR [+voice]	100	106
Mean	(93)	(99)

the voiceless and voiced environments, respectively.

Insert Figure 1 about here

Vowel duration was longer preceding underlying voiced stops than underlying voiceless stops in: (1) /rrik/-/rrig/ in the voiceless assimilatory environment (top panel) [$F(1,4) = 47.64$; $p < 0.003$] (mean difference = 22 ms) and (2) /fat/-/fad/ in the voiced assimilatory environment (bottom panel) [$F(1,4) = 30.08$; $p < 0.006$] (mean difference = 20 ms). Thus, vowel duration distinguished underlying voicing in the assimilatory environments only 20 percent of the time in the contexts where semantically biasing information is present.

For the non-semantically biasing context, a significant main effect of vowel duration was found [$F(1,4) = 53.36$; $p < 0.002$]. In addition, significant two-way interactions involving underlying voicing and environment [$F(2,8) = 37.69$; $p < 0.001$] and involving underlying voicing and minimal pair [$F(2,8) = 17.10$; $p < 0.000$] were obtained. Moreover, a three-way interaction involving underlying voicing, environment, and minimal pair was also significant [$F(8,32) = 3.93$; $p < 0.003$].

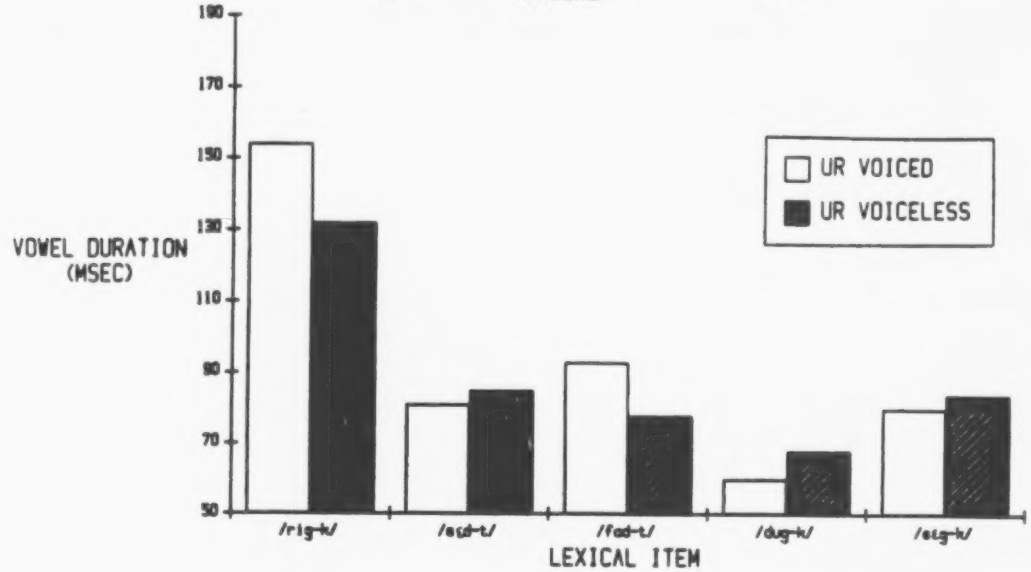
One-way analyses of variance performed on the three-way interaction revealed that the main effect of underlying voicing was attributable to three minimal pairs in both assimilatory environments. Figure 2 shows the results for the minimal pairs. The format is the same as Figure 1.

Insert Figure 2 about here

In the voiceless assimilatory environment (top panel), vowel duration was longer preceding underlying voiced stops than preceding underlying voiceless stops only in the pairs: (1) /rrik/-/rrig/ [$F(1,4) = 14.71$; $p < 0.02$], (2) /sɛt/-/sɛd/ [$F(1,4) = 81.89$; $p < 0.001$], and (3) /fat/-/fad/ [$F(1,4) = 15.04$; $p < 0.02$]. In the voiced assimilatory environment (bottom panel), vowel duration was longer preceding underlying voiced stops than underlying voiceless in the same three minimal pairs: (1) /rrik/-/rrig/ [$F(1,4) = 64.55$; $p < 0.002$], (2) /sɛt/-/sɛd/ [$F(1,4) = 25.62$; $p < 0.008$], and (3) /fat/-/fad/ [$F(1,4) = 117.95$; $p < 0.001$]. Thus, underlying voicing was distinguished 60 percent of the time in the non-semantically biasing context.

Recall that if regressive voice assimilation applies such that word-final stops assume the voicing of the following consonant, then mean vowel duration across underlying voiced and voiceless stops should be longer in the voiced assimilatory environment relative to the voiceless assimilatory environment. This prediction is based on the fact that vowel duration, as a correlate to voicing, is longer preceding word-final voiced stops than voiceless stops. Figure 3 shows the mean vowel duration results collapsed across lexical items as a function of the assimilatory environment. The top panel shows the results for the semantically biasing context and the bottom panel the results

SEMANTICALLY BIASING - [-VOICE] ENV.



SEMANTICALLY BIASING - [+VOICE] ENV.

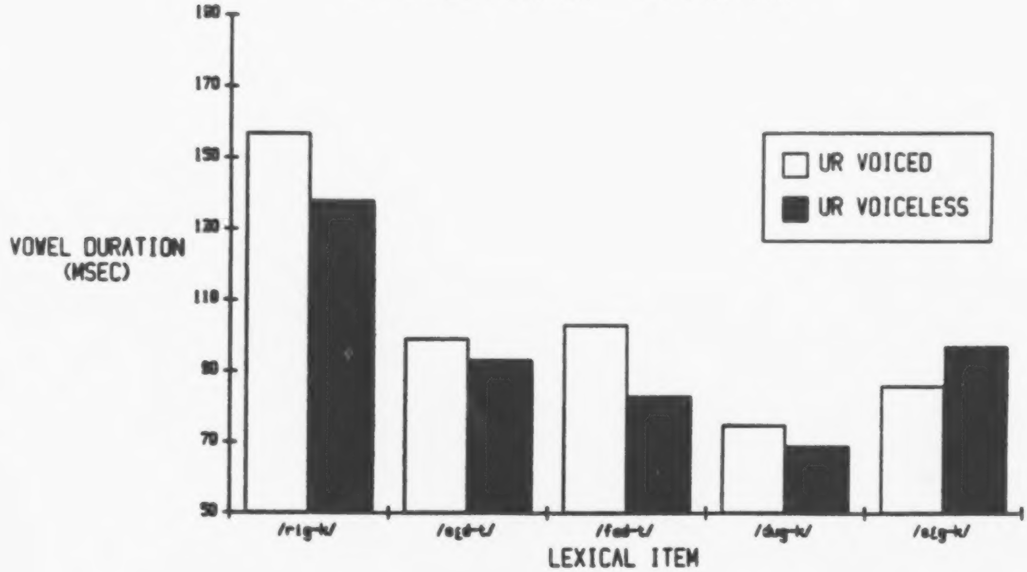


Figure 1. Vowel duration results for each minimal pair produced in semantically biasing contexts.

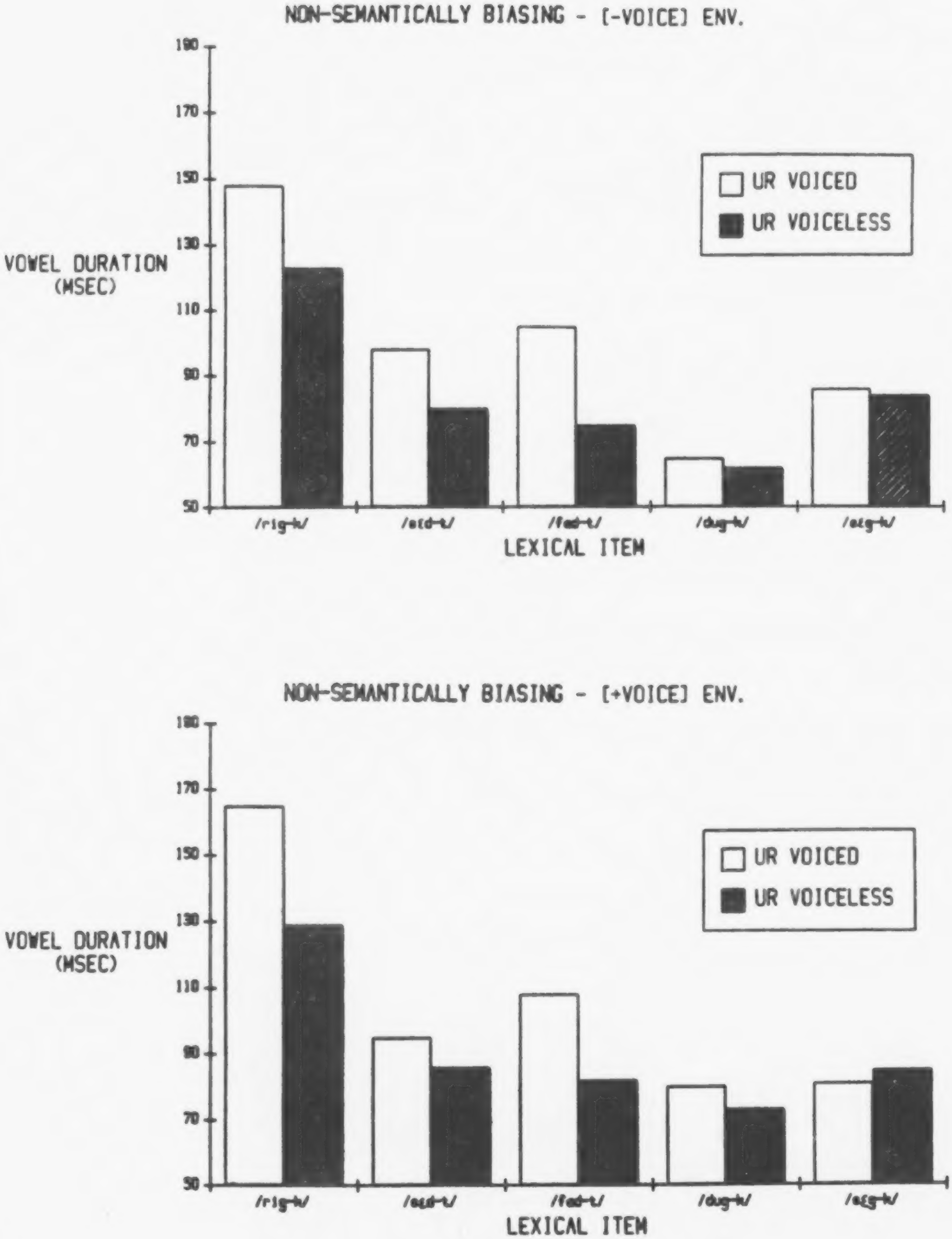


Figure 2. Vowel duration results for each minimal pair produced in non-semantically biasing contexts.

for the non-semantically biasing context.

Insert Figure 3 about here

A significant difference between assimilatory environments was obtained by *only* for the semantically biasing context [$F(1,4) = 21.77$; $p < 0.01$]. Vowel duration collapsed across underlying voicing was 9 ms longer in the voiced assimilatory environment than in the voiceless assimilatory environment. This is shown in Figure 3 by comparing the right pair of bars (mean duration = 100 ms) with the left pair of bars (mean duration = 91 ms).

There was no significant difference between assimilatory environments in the non-semantically biasing context [$F(1,4) = 6.93$; $p < 0.06$]. The mean difference across underlying voicing between the right pair of bars and the left pair is 5 ms. However, in this context type, vowel duration significantly distinguished underlying voicing [$F(1,4) = 77.24$; $p < 0.001$]. This is shown in Figure 3 by comparing the open bars with the filled bars in the right and left pairs of bars. Averaged across both environments, vowel duration was 15 ms longer preceding underlying voiced stops (mean duration = 103 ms) than preceding underlying voiceless stops (mean duration = 88 ms). This set of results suggests that when semantically biasing information is present, vowel duration shows the predicted effects of regressive voice assimilation. Thus, neutralization appears to result. However, when semantically biasing information is absent, then underlying voicing is distinguished regardless of the assimilatory environments and neutralization is incomplete.

Voicing during Closure

Table V shows the mean voicing during closure durations collapsed across lexical items. The format is the same as Table IV.

Insert Table V about here

No significant main effect of underlying voicing was found for the semantically biasing context [$F(1,4) = 2.09$; $p < 0.3$] or for the non-semantically biasing context [$F(1,4) = 3.54$; $p < 0.2$]. Moreover, there were no significant interactions involving underlying voicing for either semantic context.

The main effect of environment was significant for both types of contexts. For the semantically biasing context [$F(1,4) = 16.62$; $p < 0.002$], voicing during closure collapsed across underlying voicing was longer in the voiced assimilatory environment and shorter in the voiceless assimilatory environment. Likewise, for the non-semantically biasing context [$F(1,4) = 14.07$; $p < 0.003$], voicing during closure averaged across underlying voicing was longer in the voiced assimilatory environment and shorter in the voiceless

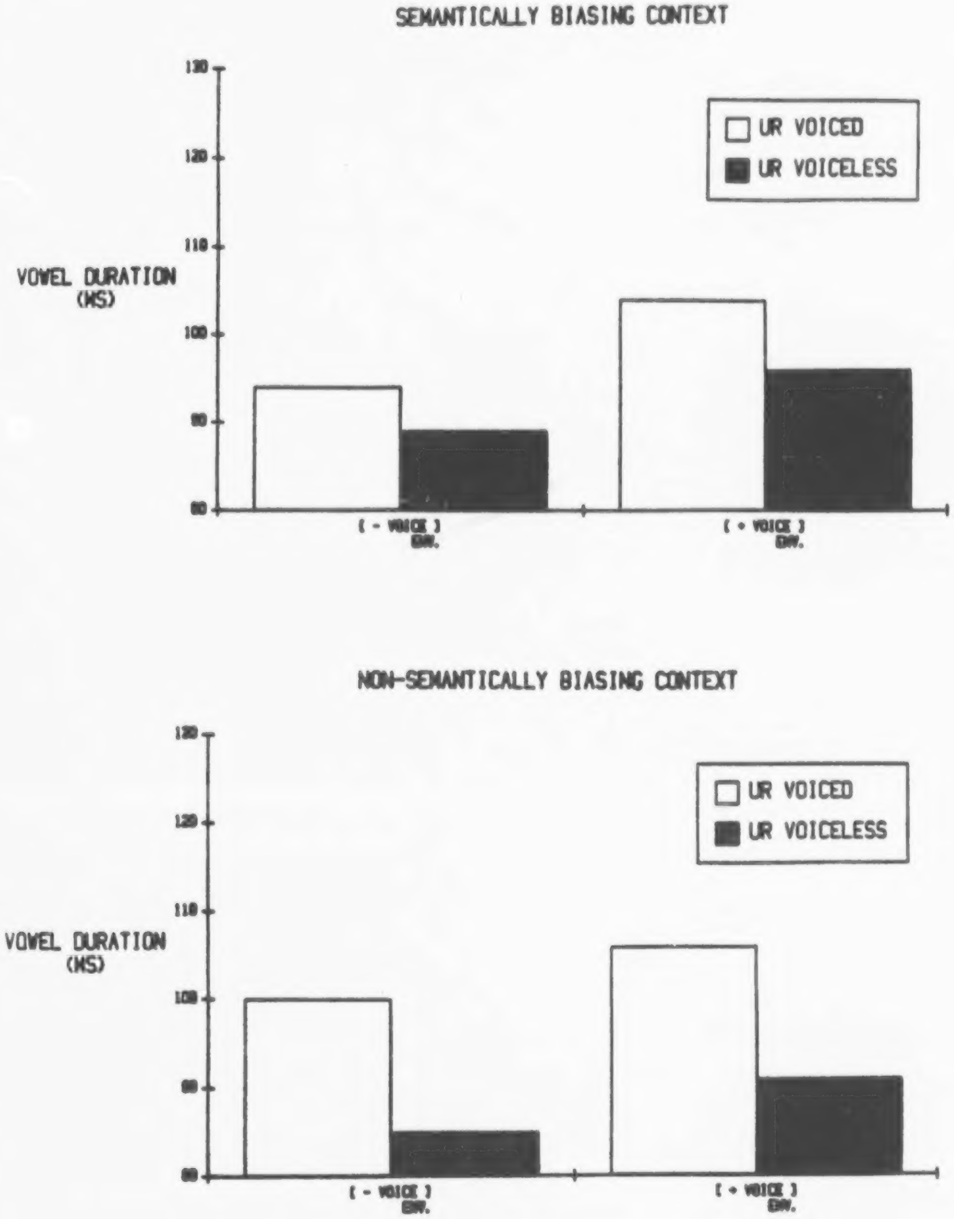


Figure 3. Vowel duration results across minimal pairs as a function of assimilatory environments and type of semantic context.

Table V

Mean voicing during closure durations (ms) for underlying voiceless and voiced stops produced in the semantically biasing context (top) and in the non-semantically biasing context (bottom) and in each of the two environments: (1) following voiceless assimilatory environment (/_ #[-voice]) and (2) following voiced assimilatory environment (/_ #[+voice]).

Voicing during Closure

	Semantically Biasing Context	
	/___#[-voice]	/___#[+voice]
UR [-voice]	18	23
UR [+voice]	20	20
Mean	(19)	(23)
	Non-Semantically Biasing Context	
UR [-voice]	19	23
UR [+voice]	18	21
Mean	(19)	(22)

assimilatory environment.

Voicing during closure appears to reflect the operation of voicing assimilation rules in both types of semantic context. In particular, voicing during closure was longer in the voiced environment relative to the voiceless environment, which is the predicted direction as a correlate of voicing.

Closure Duration

Table VI shows the mean closure durations across lexical items. Again, the format is identical to Tables IV and V.

Insert Table VI about here

No significant main effect of underlying voicing was found for the semantically biasing context [$F(1,4) = 1.25$; $p < 0.4$] or for the non-semantically biasing context [$F(1,4) = 2.34$; $p < 0.03$].

Environment was significantly different between voiceless and voiced environments only for the semantically biasing context [$F(1,4) = 10.42$; $p < 0.04$]. Environment was not significantly different for the non-semantically biasing context [$F(1,4) = 1.22$; $p < 0.4$].

If regressive voice assimilation applies to word-final stops in the semantically biasing context, then closure duration should be longer before a voiceless consonant than before a voiced consonant. Again, this prediction follows from the fact that closure duration, as a correlate of voicing, is longer for voiceless stops than for voiced stops. The opposite is found in the present results. However, it cannot be inferred that assimilation did not apply because the following consonants determining voicing differ articulatorily and acoustically in ways other than voicing. The differences in closure duration between the voiced and voiceless assimilatory environments may be a result of coarticulation and compensatory adjustment in the timing between different segment types. For example, sequential consonantal gestures are more similar between the word-final stops and the following trill (or a series of short stops) than they are between the final stops and the following fricative [cf. Lindblom, 1983]. Because the gestures required to complete the obstructions for the [-continuant] segments overlap, the result may be less precise articulatory gestures. Thus, stop closure duration may be shorter preceding a trill because of their similar articulatory gestures. However, when stops precede fricatives, more time and energy may be necessary to complete one set of gestures for the stop and begin a different set of gestures for the fricative. This may result in more precise articulations and, therefore, longer closure durations.

The fact remains, however, that, like the results for vowel duration, significant differences between assimilatory environments are found only in the semantically biasing context and not in the non-semantically biasing context. Thus, assimilation is more apparent when semantically biasing information is present to constrain the test words.

Table VI

Mean closure durations (ms) for underlying voiceless and voiced stops produced in the semantically biasing context (top) and in the non-semantically biasing context (bottom) and in each of the two environments: (1) following voiceless assimilatory environment (/_ #[-voice]) and (2) following voiced assimilatory environment (/_ #[+voice]).

Closure Duration

	Semantically Biasing Context	
	/___#[-voice]	/___#[+voice]
UR [-voice]	76	99
UR [+voice]	76	97
Mean	(76)	(98)
	Non-Semantically Biasing Context	
UR [-voice]	66	76
UR [+voice]	84	88
Mean	(75)	(82)

Summary and Discussion

Of the three temporal intervals measured, only vowel duration distinguished underlying voicing. Table VII summarizes the results.

Insert Table VII about here

Vowel duration distinguished voicing only in the non-semantically biasing context. In addition, the results from vowel duration (and less conclusively closure duration) showed that the assimilatory environments were different only in the semantically biasing context but not in the non-semantically biasing context. This suggests that when semantically biasing information is lacking, underlying voicing is distinguished, thereby blocking the application of the assimilation rule. However, when semantic information is present, underlying voicing is not distinguished and the assimilation rule appears to apply. This results in voiced stops in a following voiced assimilatory environment or voiceless stops in a following voiceless assimilatory environment. Thus, the results from this investigation showed that the degree of semantically biasing information in an utterance can affect the neutralization of an underlying voice contrast.

Syntax was always present to constrain the occurrence of a test word in both the semantically biasing and non-semantically biasing contexts. The important difference between the contexts, then, is the fact that semantic information was present or absent to constrain and bias the meaning of the test word. The results suggest that, without semantically redundant information, speakers may more readily distinguish the underlying voice contrast. Syntactic information does not appear to effect voicing neutralization to the extent that semantic information does. This is not to say that syntax does not interact with neutralization processes. If syntax were not present to constrain the lexical choice of a target word, then there should be a 50 percent chance that neutralization would result for each of the five speakers when semantically biasing information is lacking. Stated otherwise, without some constraint in the non-semantically biasing context, speakers can choose to produce either member of a given minimal pair. Syntax does play some role to constrain lexical choice and, in the non-semantically biasing context, this choice is phonetically realized in the duration of the vowel preceding either a voiced or voiceless stop. However, semantic information appears to override syntactic information, as evidenced by the results from the semantically biasing context. When semantic information was available to bias the intended meaning, underlying voicing was not distinguished and the minimal pairs were essentially homonyms. The presence of semantically biasing (and redundant) information afforded the process of neutralization to be complete, at least with respect to the temporal measurements made.

The overall results demonstrate that there is an interaction between semantic information and the phonological neutralization phenomenon of regressive voice assimilation. When semantically biasing information is lacking, underlying voicing is distinguished and voicing assimilation is precluded. When semantically biasing information is present, the underlying word can be recovered through context. In this situation, then, voicing assimilation can occur, thereby obliterating the underlying contrast that

Table VII

Summary of the effects of semantic context on regressive voice assimilation. The results are summarized for the semantically biasing context and the non-semantically biasing context in the top and the bottom panels, respectively. Each row summarizes the results as a function of the environment in which test words were produced

Semantically Biasing Context

	Environment	Vowel Duration	Voicing During Closure	Closure Duration
1. Underlying Voicing Distingished	[-voice]	no	no	no
	[+voice]	no	no	no
2. Evidence of Assimilation	[-voice]	yes	yes	yes
	[+voice]	yes	yes	yes

Non-Semantically Biasing Context

	Environment	Vowel Duration	Voicing During Closure	Closure Duration
1. Underlying Voicing Distingished	[-voice]	yes	no	no
	[+voice]	yes	no	no
2. Evidence of Assimilation	[-voice]	no	yes	no
	[+voice]	no	yes	no

distinguishes between words, without possible ambiguity resulting.

A possible explanation as to why production differences were found to distinguish underlying voicing in the non-semantically biasing contexts and not in the semantically biasing contexts may lie in how the speakers imposed emphatic, or contrastive, stress [e.g., Bolinger, 1961; Chafe, 1974]. For example, in the semantically biasing context, if the target word duke had already been semantically primed, for example, then it is old information for both speaker and listener. On the other hand, in the non-semantically biasing context, duke would not have been primed previously, or otherwise activated, and thus it would constitute new information for the speaker. Consequently, the speaker may place more stress, relative to other words in the utterance, on this word to indicate its status as new information. This stress would be phonetically realized, among other parameters, by lengthening of the stressed vowel, as well as overall lengthening of the word and more precise articulation. These acoustic manifestations could give rise to the underlying representations being distinguished in the non-semantically biasing context. Although old, non-stressed information versus new, stressed information may provide the explanation for the present results, the presence or absence of semantically biasing information is still responsible for establishing what is new and what is old, at least in the mind of the speaker. (See Fowler and Housum [1987] for an excellent demonstration of speakers' use and listeners' perception of old and new information.)

Thus, it remains that speakers must assess old and new information to determine when to apply stress. Furthermore, one result of stress is greater intelligibility of the speech signal [Lieberman, 1963]. In general, greater intelligibility is the result of less reduced speech and, therefore, less coarticulation. Perhaps, then, a more encompassing explanation of why assimilation was found only in the semantically biasing context involves coarticulatory processess. Phonological assimilations are abstract descriptions of coarticulation [see Lindblom, 1983]. Effects of coarticulation are generally strong in fluent speech, even across word-boundaries [Oshika, Zue, Weeks, Neu, and Aurbach, 1975; cf. Church, 1987]. However, the effects of coarticulation may be weakened in contexts that lack higher levels of linguistic information. Thus, in the present study, the effects of coarticulation of voicing may be weakened because semantically biasing information is lacking. In this context, the minimal pairs are more clearly, or precisely, articulated, presumably to facilitate accessing the correct underlying morpheme [Lieberman, 1963; cf. Hunnicutt, 1985]. This suggests that certain coarticulatory processes may be precluded depending upon the whole semantic construct. Without the higher level information to ensure that the gist of the utterance is communicated, then bottom-up, acoustic-phonetic information becomes more important in the communicative exchange.

Thus, durational differences distinguishing underlying voicing may be the result of new information and resulting stress or the result of a more general weakening of coarticulatory processes when semantically biasing information is lacking. The conditioning factor, however, remains the presence and absence of semantically biasing information and the consequential acoustic-phonetic effect of complete or incomplete neutralization. This suggests, then, that there may be some on-line assessment of the degree of semantically biasing information by speakers. The result is an assignment of a semantic weight to an utterance. These weights are established by summing across the degree of semantically biasing (or redundant) information present in an utterance to constrain the meaning of a word meeting the structural description of, for example, regressive voice assimilation. The application of voice assimilation (and presumably certain other phonological processes--e.g., post-lexical) is sensitive to or conditioned by these weights. Thus, when a high degree of semantically biasing information is lacking to constrain meaning, voicing assimilation is blocked and the underlying voice contrast is phonetically realized. However, when a high degree of semantically biasing information is present, voice assimilation applies and the underlying contrast is neutralized. It stands to reason that these weights are necessarily gradient, or continuous, in nature, dependent upon the individual speaker's assessment of the degree of semantically biasing information. As a result, individual differences as to complete versus incomplete neutralization may be observed.

The introduction of semantic weights account in a psychologically real way for the role that the semantics, arising from sentence formation and lexical insertion, have in constraining phonological processes. Furthermore, they account for the empirical findings demonstrating that neutralization processes are not independent of other aspects of linguistic knowledge, especially a speaker's semantic interpretation or assessment of semantically biasing information of a given utterance.

In conclusion, the present study demonstrates that phonological neutralization processes are indeed affected by the presence and absence of semantically biasing information. Phonology does not occur in a vacuum. Phonological processes are part of an interactive linguistic system. In order to arrive at a consistent and systematic explanation of the kinds of phonological phenomenon examined in this study, higher levels of linguistic information must be taken into account.

Footnotes

1 Note that the two neutralization processes of word-final devoicing and regressive voice assimilation overlap in the voiceless assimilatory environment [DeCesaris, 1980]. However, the putative phonetic result is the same. Both underlying voiced and voiceless stops are realized as voiceless.

2 Subjects were also asked to produce the target minimal pairs in an utterance final environment in each of the semantic contexts. This is the ideal word-final devoicing environment, where underlying voiced and voiceless stops putatively become voiceless [Wheeler, 1979]. In the utterance final devoicing environment neutralization appears to be complete, regardless of the type of context. However, as a consequence of utterance final lengthening, any durational differences that might be present to distinguish the underlying voice contrast in a non-semantically biasing context appear to be superceded by the syntactically imposed durational modification. For ease of exposition, I will not report in the present report the specific results from the utterance final environment. (For a complete discussion see Charles-Luce [1987].)

References

Aaronson, D.; Scarborough, H. S.: Performance theories for sentence coding: Some quantitative evidence. J. Exp. Psych.: Hum. Percep. Perform. 2: 56-70 (1976).

Bolinger, D. L.: Contrastive accent and contrastive stress. Lang. 37: 83-96 (1961).

Bransford, J. D.; Franks, J. J.: The abstraction of linguistic ideas. Cog. Psych. 2: 331-350 (1971).

Bransford, J. D.; Johnson, M. K.: Contextual prerequisites for understanding: Some investigations of comprehension and recall. J. Verb. Learn. Verb. Behav. 11: 717-726 (1972).

Chafe, W. L.: Language and consciousness. Lang. 50: 111-133 (1974).

Charles-Luce, J.: Word-final devoicing in German: Effects of phonetic and sentential contexts. J. Phonet. 13: 309-324 (1985).

Charles-Luce, J.: An acoustic investigation of neutralization in Catalan. (Unpublished dissertation, Indiana University, Bloomington, Ind. 1987).

Charles-Luce, J.; Walker, L. A.: Effects of linguistic context on the durations of lexical categories; in Research on Speech Perception: Progress Report No. 7 (Speech Research Laboratory, Indiana University, Bloomington, Ind. 1981).

Chen, M.: Vowel length variation as a function of the voicing of the consonant environment. Phonet. 22: 129-159 (1970).

Church, K. W.: Phonological parsing and lexical retrieval; in Frauenfelder, Tyler, Spoken word recognition, pp. 53-69 (MIT Press, Cambridge 1987).

DeCesaris, J.: Consonant alternations in Catalan. Innovat. Ling. Ed. 1: 65-84 (1980).

Dinnsen, D. A.; Charles-Luce, J.: Phonological neutralization, phonetic implementation and individual differences. J. Phonet. 12: 49-60 (1984).

Dooling, D. J.; Lachman, R.: Effects of comprehension on retention of prose. J. Exp. Psych. 88: 216-222 (1971).

Fourakis, M.; Iverson, G. K.: On the "incomplete neutralization" of German final obstruents. Phonetica 41: 140-149 (1984).

Fowler, C. A.; Housum, J.: Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. J. of Mem. and Lang. 26: 489-504 (1987).

House, A. S.; Fairbanks, G.: The influence of consonantal environment upon the secondary acoustical characteristics of vowels. J. Acoust. Soc. Am. 25: 105-113 (1953).

Hunnicutt, S.: Intelligibility versus redundancy--conditions of dependency. Lang. and Sp. 28: 47-56 (1985).

Lieberman, P.: Some effects of semantic and grammatical context on the production and perception of speech. Lang. Sp. 6: 172-187 (1963).

Lindblom, B.: Economy of speech gestures; in Peter F. MacNeilage, The production of speech, pp. 217-245 (Springer-Verlag, New York 1983).

Luce, P. A.; Carrell, T. D.: Creating and editing waveforms using WAVES; in Research on Speech Perception: Progress Report No. 7, (Speech Research Laboratory, Indiana University, Bloomington, Ind. 1981).

Mascaro, J.: Catalan Phonology and the Phonological Cycle. (Indiana University Linguistics Club, Bloomington, Ind. 1978.)

Miller, G. A.; Heise, G. A.; and Lichten, W.: The intelligibility of speech as a function of the context of the test materials. J. Exp. Psych. 41: 329-335 (1950).

Miller, G. A.; and Isard, S.: Some perceptual consequences of linguistic rules. J. Verb. Learn. Verb. Behav. 2: 217-228 (1963).

Moulton, W.: The Sounds of English and German (University of Chicago Press, Chicago, 1962).

Ohala, J. J.: The origin of sound patterns in vocal tract constraints; in Peter F. MacNeilage, The Production of Speech (Springer-Verlag, New York, 1983).

Oshika, B. T.; Zue, V. W.; Weeks, R. V.; Neu, H.; Aurbach, J.: The role of phonological rules in speech understanding research. IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-23, 104-112 (1975).

Pollack, I.; Pickett, J. M.: The intelligibility of excerpts from conversation. Lang. Sp. 6: 165-171 (1963).

Pollack, I.; Pickett, J. M.: Intelligibility of excerpts from fluent speech: auditory vs. structural context. J. Verb. Learn. Verb. Behav. 3: 79-84 (1964).

Port, R.; O'Dell, M.: Neutralization of syllable final voicing in German. J. Phonet. 13: 455-471 (1985).

Slowiaczek, L.; Dinnsen, D. A.: On the neutralizing status of Polish word-final devoicing. J. Phonet. 13: 325-341 (1985).

Wheeler, M.: Phonology of Catalan (Blackwell, Oxford, 1979).

Stimulus Variability and Processing Dependencies in Speech Perception*

John W. Mullennix and David B. Pisoni

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, IN 47405

---

*The research reported here was supported by NIH Research Grant NS-12179-11 and NIH Training Grant NS07134-09 to Indiana University in Bloomington, IN. The authors would like to thank Luis Hernandez for programming assistance.

Abstract

Processing dependencies in speech perception between voice and phoneme were investigated using the Garner (1974) speeded-classification procedure. Variability in the voice of the talker and in the cues to word-initial consonants were manipulated and their effects on performance observed. The results showed that the processing of a talker's voice and the perception of voicing were asymmetrically integral. In addition, when stimulus variability was increased in each dimension, the amount of orthogonal interference obtained for each dimension became significantly larger. The processing asymmetry between voice and phoneme was interpreted in terms of a parallel-contingent relationship of talker normalization processes to auditory-to-phonetic coding processes in speech perception. The effects of talker variability provided additional evidence showing that variation from trial-to-trial in the voice of the talker results in reliable and robust effects on speech perception and spoken word recognition. Effects of talker variability do not appear to be independent or dissociated from the encoding of the phonetic information in the speech signal.

Stimulus Variability and Processing Dependencies in Speech Perception

The production of human speech is characterized by a large number of individual differences between talkers. Such factors as structural differences in vocal tract size and shape (Fant, 1973; Joos, 1948; Peterson & Barney, 1952), glottal characteristics (Carr & Trill, 1964; Carrell, 1984; Monsen & Engebretson, 1977), and dynamic articulatory control (Ladefoged, 1980), etc. manifest themselves in the speech waveform in terms of a variety of acoustic differences between talkers. One of the major issues in speech perception concerns the manner in which the acoustic differences between talkers are processed in perceiving spoken language. It is likely that several processes and/or mechanisms exist that perform some type of perceptual compensation on talker voice information in order to facilitate the extraction of linguistic units germane to speech. Some researchers have characterized these processes as "normalizing" or "adjusting" the acoustic differences between talkers (e.g. Summerfield, 1975; Summerfield & Haggard, 1973). However, the manner in which these processes operate has not been clearly described and a precise characterization of such processes has not been developed. Although some research has been devoted to this problem (see below), for the most part the perceptual consequences of these compensation processes have not been fully investigated. Most studies in speech perception over the last forty years have used speech produced by one talker. And, frequently only one token of each utterance is used as the stimulus material, therefore preventing any systematic assessment of the role of stimulus variability in perception.

With regard to the perceptual consequences of processing the acoustic differences between talkers, experimental research examining vowel and consonant perception (Assman, Nearey, & Hogan, 1982; Fourcin, 1968; Rand, 1971; Verbrugge, Strange, Shankweiler, & Edman, 1976; Weenink, 1986), word recognition (Creelman, 1957; Mullennix, Pisoni, & Martin, 1987), and memory (Martin, Mullennix, Pisoni, & Summers, 1987) has demonstrated that changes in the voice of the talker from trial to trial within an experiment result in a decrement in overall task performance. The presence of these effects can be interpreted in terms of a "processing cost" to the perceptual system that is induced by variability in the talker's voice. For instance, in one recent study, Mullennix, Pisoni, and Martin (1987) examined the effects of talker variability on spoken word recognition. In a number of experiments using perceptual identification and word naming tasks, we et al. found that word recognition was significantly worse for words produced by different talkers compared to the same words produced by only a single talker. Furthermore, we observed that when the acoustic information in the speech signal became increasingly degraded by using a special distortion technique, the effects of talker variability on performance became even greater. Because perceptual performance was consistently worse when the words were produced by different talkers, we suggested that a resource-demanding perceptual mechanism is probably employed by listeners to compensate for the acoustic differences. In addition, because these effects were greater when the early acoustic information in the signal was disrupted, we suggested that the processing of voice information is closely related to processes involved in the early perceptual encoding of the input signal into an initial phonetic representation. Our results provided the first step to characterizing the nature of talker-related perceptual processes. However, the relationship of these processes to other phonetic coding processes and to the higher-level processes involved in word recognition and lexical access are largely unknown and remain a topic for additional investigation.

One important aspect of "talker normalization" processes is concerned with the relationship of these processes to the auditory-to-phonetic coding processes of speech. Do the perceptual processes used to encode voice information function independently of processes that are used to encode phonetic information in the speech signal? Or, are talker normalization processes and phonetic coding processes interrelated? A major objective of the present study was to investigate the relationship of talker normalization processes and auditory-to-phonetic coding processes and assess their interactions. One way to determine whether perceptual processes are related to one another is to assess whether stimulus dimensions relevant to both types of processes are perceived independently of one another or whether there is some dependency relation. In the present study, we examined the nature of the processing relations between talker normalization and auditory-to-phonetic coding processes by using an experimental technique specifically designed to study processing interactions between two stimulus dimensions (see Garner, 1974).

As mentioned earlier, one hypothesis that has been proposed to account for talker variability effects is that a resource-demanding perceptual mechanism that processes talker voice information is invoked each time a word is presented in a different voice (Mullennix et al., 1987). That is, talker normalization processes that require limited-capacity processing resources to perform their operations are engaged to encode the voice of the talker. According to this account, perceptual deficits due to changes in a talker's voice occur because of competition for processing resources used by talker normalization processes and other perceptual processes used in speech perception. Closely related to this is the issue of the controllability of these processes. It is conceivable that each time a different voice is encountered, control processes give temporary priority to talker normalization processes until voice-related perceptual operations are completed. If this is the case, perceptual deficits may arise from the additional time it takes to switch control back and forth between talker normalization processes and other perceptual processes used to construct a phonetic representation from the speech signal. If the allocation of processing resources to both types of processes is related to selective attention, or, if selective attention to speech-related processes is affected by shifts of processing control to talker-related processes, then the effects of talker variability may be intimately dependent on the role of selective attention in speech perception. By examining the processing interactions between word-related and talker-related stimulus dimensions, we hoped to obtain further information about the role of selective attention in speech perception and spoken word recognition and assess the interactions of these dimensions.

Another issue addressed in the present investigation concerns the effects of trial-to-trial stimulus variability in speech perception. Previous studies using perceptual indentification and naming tasks have found that trial-to-trial variability in the voice of the talker resulted in significant decrements in word recognition (i.e. Creelman, 1957; Mullennix et al., 1987). That is, the acoustic variation in the words produced by different talkers led to poorer recognition performance. In the present study, the voice of the talker and the acoustic-phonetic composition of word-initial consonants were manipulated in a speeded-classification task. If trial-to-trial changes in variability have detrimental effects on performance using this task, the results would provide additional evidence that stimulus variability from trial to trial produces significant perceptual effects on spoken word recognition. By manipulating word variability and talker variability together, we hoped to obtain further information about the potential interactions of these two variables.

In order to examine the nature of any processing dependencies between talker normalization and auditory-to-phonetic coding processes, and, to assess the extent to which talker normalization processes are related to selective attention, a modified version of the selective attention procedure described by Garner (1974) was used. Over the years, this procedure has been adopted by a number of researchers to examine processing dependencies between auditory and phonetic dimensions (Blechner, Day, & Cutting, 1976; Carrell, Smith, & Pisoni, 1981; Eimas, Tartter, Miller, & Keuthen, 1978; Miller, 1978; Pastore et al., 1976; Tomiak, Mullennix, & Sawusch, 1987; Wood, 1974; Wood & Day, 1975). These studies have shown that certain stimulus dimensions relevant to speech are processed as integral dimensions, often displaying a mutual dependence on each other.

The experimental procedure developed by Garner (1974) involves a two-choice speeded classification task. Subjects are required to selectively attend to one stimulus dimension while simultaneously ignoring another stimulus dimension. Two stimulus dimensions are combined in various ways to form three types of stimulus sets: A control set, an orthogonal set, and a correlated set. In the control set, the unattended dimension is held constant while the attended dimension varies randomly. The control set for each dimension provides a baseline measure for classifying each dimension and permits one to assess whether both dimensions are equally discriminable. In the orthogonal set, both the attended and unattended dimensions vary randomly. The degree to which response latencies increase from the control set to the orthogonal set for each dimension indicates the extent to which the stimulus dimensions are processed separably or in an integral fashion. If stimulus dimensions are classified as quickly in the orthogonal conditions as they are in the control conditions, then the stimulus dimensions are said to be processed independently. In this case, filtering out the irrelevant dimension is relatively complete. However, if there is a significant increase in response latencies from the control conditions to the orthogonal conditions, the stimulus dimensions are said to be processed in a dependent manner. That is, the variation in the irrelevant dimension cannot be selectively ignored or filtered by the subject and the processing of the irrelevant dimension interferes with the processing of the attended dimension. This result is termed "orthogonal interference" and it indicates that a failure of selective attention to the attended dimension has occurred. Finally, in the correlated condition, one particular value of one dimension is always paired with another particular value of the other dimension. The presence of decreased response latencies in this condition compared to the control condition is called a redundancy gain. A redundancy gain indicates that the information in the non-attended stimulus dimension can be used to facilitate perceptual classification. Although the presence of a redundancy gain can be interpreted as further evidence for integrality of dimensions (see Garner, 1974; Garner & Felfoldy, 1970), it is best thought of as additional evidence and is not crucial for making assertions about integral processing. However, under certain circumstances, the presence of redundancy gains can provide important evidence regarding the serial/parallel nature of the processes involved (Wood, 1974, 1975) or it can reveal the presence of a selective serial processing strategy (Biederman & Checkosky, 1970; Felfoldy & Garner, 1971).

In the present study, the processing relationship between talker normalization and phonetic coding was examined by manipulating one stimulus dimension related to the talker's voice and one stimulus dimension related to phonetic categorization. To avoid confusion, the two stimulus dimensions selected were called the "voice" dimension and the "word" dimension. The voice dimension involved variations in the gender of the talker (i.e., male versus female). The word dimension involved variations in the phonetic feature of voicing (/b/ versus /p/) in initial position. When subjects were required to attend to the voice dimension, the required responses were "male voice" or "female voice"; when the subjects were required to attend to the word dimension, the required responses were "b" or "p". By examining performance in classifying these two dimensions using the selective attention procedure, we hoped to assess the degree of separability and/or integrality of the two stimulus dimensions.

The second manipulation we were concerned with was related to stimulus variability in speech perception. Word variability and talker variability were manipulated together by changing the composition of the orthogonal stimulus set. Word variability was increased by increasing the number of "b" and "p" words within the orthogonal set so that the acoustic-phonetic composition of word-initial consonants could be varied across the words. Talker variability was increased by increasing the number of male and female talkers producing the words used within the orthogonal set. By comparing the amount of orthogonal interference obtained across conditions, the effects of stimulus variability in word and voice information on speeded classification of these two dimensions could be assessed.

A number of predictions concerning the outcome of the present experiment can be made. First, we consider the response latencies in the control and orthogonal conditions. If there is no increase in response latencies from the control condition to the orthogonal condition for either the word dimension or the voice dimension, this pattern of results would suggest that the two dimensions are processed in a separable manner. This outcome would be consistent with the idea that the encoding of voice information is carried out by processes that function independently of the processes used to extract phonetic information from the speech signal. This result would also suggest that the effects of talker variability found in previous studies are probably not due to a failure of selective attention to the phonetically-relevant acoustic information contained in the word. However, if there are significant increases in response latencies from control to orthogonal conditions for both stimulus dimensions, this pattern of results would suggest that the processing of voice information and the perception of voicing are integral. These results would also imply that auditory-to-phonetic coding processes and talker normalization processes are highly interrelated. If redundancy gains are obtained for either dimension, this would provide further evidence of integrality and would permit one to conclude that the two processes operate in parallel. The presence of integrality effects in these conditions would also provide support for the assertion that the processing of voice information is mandatory and that the effects of talker variability observed in previous studies is probably related to a failure of selective attention to phonetically-relevant information in the speech signal.

Our final prediction concerns the effects of stimulus variability on speeded classification. If the amount of orthogonal interference for either dimension becomes greater as variability in the orthogonal set is increased, this pattern of results would provide additional evidence that lexical variability and talker variability produce detrimental perceptual effects on word recognition. However, if there is no difference in the amount of interference between these conditions, this would suggest that variability does not affect performance in a task involving selective attention to stimulus dimensions. Overall, the manipulations included in the present study should provide important new information about talker normalization processes and phonetic coding processes in speech and their relationship to one another and the effects of stimulus variability in speech perception.

Method

Subjects. Seventy-two undergraduate students enrolled in introductory psychology courses at Indiana University volunteered to be subjects. Each subject took part in one 1-hour session and received partial course credit for participating in the experiment. All subjects were native speakers of English and reported no history of a speech or hearing disorder at the time of testing.

Stimulus Materials. The stimuli consisted of 16 naturally spoken words obtained from eight male and eight female talkers all of whom spoke with a midwestern dialect. The stimuli were English monosyllablic words selected from the corpus of words used in the Modified Rhyme Test (House et al., 1965). One-half of the words began with a "b" consonant and one-half of the words began with a "p" consonant. Each talker's utterances were recorded on audiotape in a sound-attenuated booth (IAC Model 401A) using an Electro-Voice Model D054 microphone and a Crown 800 series tape recorder. Each stimulus item was pronounced in citation format in unique randomized lists for each talker. The words were subsequently converted to digital form via a 12-bit analog-to-digital converter at a 10 kHz sampling rate and stored as digital files. The target words were digitally edited to produce the final experimental materials used in the study. RMS amplitude levels among words were digitally equated using a software package designed to modify digital waveforms.

Procedure. Three experimental factors were manipulated: Stimulus dimension, stimulus set condition, and stimulus variability. Stimulus dimension was manipulated within subjects by requiring subjects to attend either to the word dimension or to the voice dimension when they classified each stimulus item. Stimulus set condition was manipulated within subjects by presenting the stimuli in a control set, an orthogonal set, or a correlated set. Stimulus variability was manipulated between subjects by modifying the composition of the orthogonal stimulus sets to create four experimental conditions. In the 2W x 2T condition, the orthogonal set contained two words spoken by two talkers. In the 4W x 4T condition, the orthogonal set contained four words spoken by four talkers. In the 8W x 8T condition, the orthogonal set contained eight words spoken by eight talkers. And, in the 16W x 16T condition, the orthogonal set contained 16 words spoken by 16 talkers. Thus, as the number of different words and the number of different talkers used in the orthogonal set increased, stimulus variability increased accordingly.

The subjects were divided equally into groups and randomly assigned to the four experimental conditions. The experimental procedure used the two-choice speeded classification task developed by Garner (1974). Depending on the particular condition, subjects were required to attend to either the word dimension or the voice dimension in order to make a response. For the word dimension, subjects classified the word beginning with either an initial "b" or "p" consonant. For the voice dimension, subjects classified the word as to whether it was spoken by a male or a female.

Insert Figure 1 about here

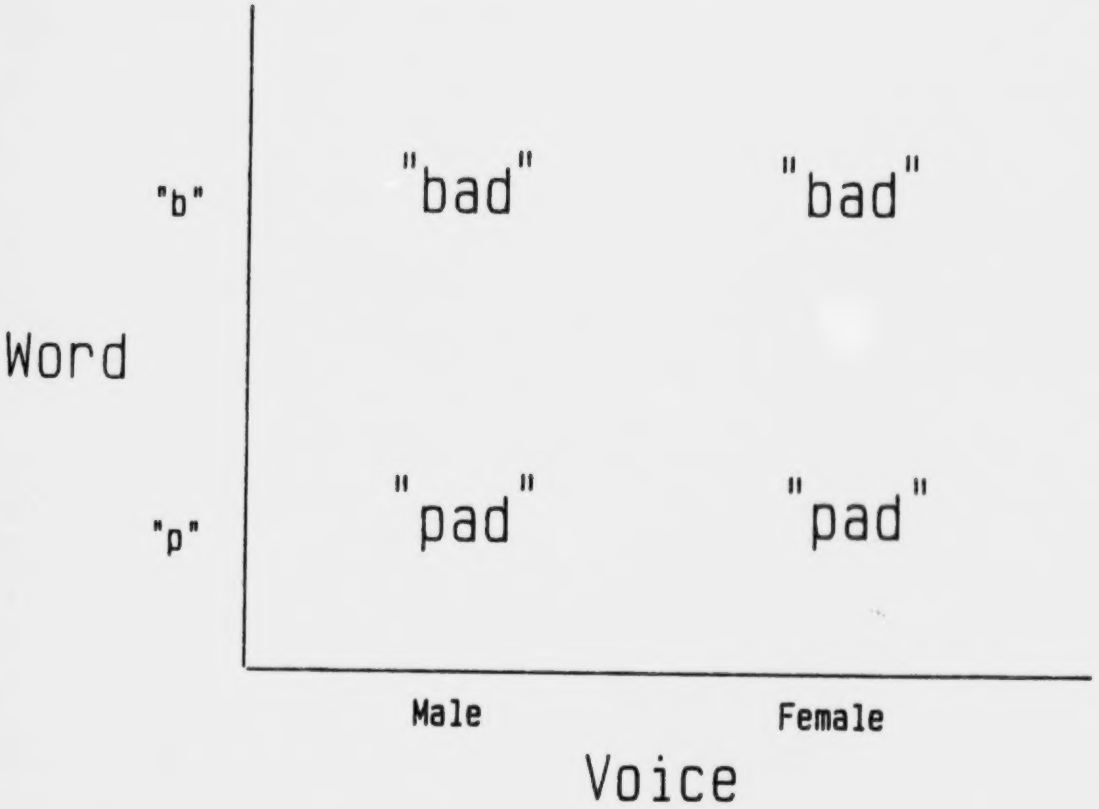

"b"
"bad"
"bad"
Word
"p"
"pad"
"pad"
Male
Female
Voice

Figure 1. The word stimuli used for the 2W x 2T condition. The stimuli are shown as a function of word dimension and voice dimension.

In Figure 1, the stimuli used for the 2W x 2T condition are displayed as a function of the word and voice dimensions. For each of the four stimulus variability conditions, subjects received three sets of trials: Control trials, correlated trials, and orthogonal trials. Thus, each subject received three sets of trials in which they classified stimuli on the word dimension and three sets of trials in which they classified stimuli on the voice dimension. In all of the control conditions, the attended stimulus dimension was varied while the irrelevant dimension was held constant. For example, one control set for the word dimension consisted of words "bad" and "pad" spoken in a male voice, while the other control set for the word dimension consisted of the words "bad" and "pad" spoken in a female voice. Each control set always contained two stimuli only. In the correlated conditions, the target dimension was always correlated with a unique irrelevant dimension. For example, one correlated set consisted of "bad" in the male voice and "pad" in the female voice, while the other correlated set consisted of "bad" in a female voice and "pad" in a male voice. The correlated sets also contained only two stimuli. In the orthogonal conditions, the stimulus dimensions varied independently. In these sets, all "b" and "p" words were presented in both male and female voices. The composition of the orthogonal sets varied across the four stimulus variability conditions.

The stimuli used in the control and correlated sets across all stimulus variability conditions were identical. These stimulus sets were formed by selecting the appropriate stimuli for each set from the words "bad" and "pad" spoken by one male talker and one female talker. However, the stimuli used in the orthogonal sets differed across the stimulus variability conditions. Table 1 shows the stimuli used for the orthogonal sets in each condition.

Insert Table 1 about here

Subjects received a total of six stimulus sets per session. The control, correlated, and orthogonal conditions were presented once for the voice dimension and once again for the word dimension. Subjects classified the first three sets in each session for one stimulus dimension and then classified the last three sets for the other stimulus dimension. The order of dimensions was counterbalanced across subjects and the order of stimulus sets was counterbalanced by means of a Latin square design. Half of the subjects received a word dimension control condition consisting of the words "bad" and "pad" spoken in a male voice and half of the subjects received a word dimension control condition consisting of the words "bad" and "pad" spoken in a female voice. In addition, half of the subjects received a voice dimension control condition consisting of the word "bad" spoken in male and female voices and half of the subjects received a voice dimension control condition consisting of the word "pad" spoken in male and female voices.

Within each stimulus set, 64 randomized test trials occurred. For the control and correlated sets, 32 repetitions of two stimuli were used. For the orthogonal sets, 16 repetitions of each stimulus occurred in the 2W x 2T condition, 4 repetitions of each stimulus in the 4W x 4T condition, and one repetition of each stimulus in the 8W x 8T and 16W x 16T conditions. Before each set of test trials, a set of 12 practice trials was presented to familiarize subjects with the specific condition. The 12 practice trials consisted of 12 stimulus items randomly selected from the set of test trials

Table 1

The list of words used in the orthogonal stimulus sets for each stimulus variability condition as a function of talker. The particular talkers are denoted by a talker number corresponding to one of the eight male talkers or one of the eight female talkers under their respective categories.

Condition	Word	Male Talker #	Female Talker #
2W x 2T	bad	1	1
	pad	1	1
4W x 4T	bad	1,2	1,2
	buff	1,2	1,2
	pad	1,2	1,2
	puff	1,2	1,2
8W x 8T	bad	1,2,3,4	1,2,3,4
	buff	1,2,3,4	1,2,3,4
	beach	1,2,3,4	1,2,3,4
	bill	1,2,3,4	1,2,3,4
	pad	1,2,3,4	1,2,3,4
	puff	1,2,3,4	1,2,3,4
	peach	1,2,3,4	1,2,3,4
	pill	1,2,3,4	1,2,3,4
16W x 16T	bad	1,2	3,4
	buff	2,3	4,5
	beach	3,4	5,6
	bill	4,5	6,7
	back	5,6	7,8
	beak	6,7	8,1
	bit	7,8	1,2
	buck	8,1	2,3
	pad	3,4	1,2
	puff	4,5	2,3
	peach	5,6	3,4
	pill	6,7	4,5
	pack	7,8	5,6
	peak	8,1	6,7
	pit	1,2	7,8
	pun	2,3	8,1

subsequently presented, with six items drawn from each response category.

The stimuli were presented binaurally over matched and calibrated TDH-39 headphones to the subject at a listening level of 80 dB. Subjects were run in small groups in sound-treated booths containing headphones and two-button response boxes. Subjects were instructed to respond as quickly and as accurately as possible by pushing one of two buttons on a computer-controlled response box in front of them. A warning light was illuminated before the presentation of each stimulus. For the practice trials, after all subjects made a response they were given feedback about the correct alternative for the trial by means of a light flashing above the button corresponding to the correct choice. Subjects did not receive feedback during the test trials. Presentation of each stimulus occurred three seconds after all subjects had made a response or three seconds after a 2-second response interval had elapsed. A 15-second interval occurred between each practice set and the appropriate test set. A one-minute rest period was inserted after each test set. Stimulus-to-response button assignment was counterbalanced across subjects. Identification accuracy and response latencies were recorded for all trials. Responses over 2000 msec were scored as incorrect and eliminated from subsequent analysis. Response latencies were measured from stimulus onset. Stimulus presentation and data collection were controlled on-line by a PDP-11/34A computer.

Results

The data were analyzed in terms of overall percent correct identification and response latencies. For each subject, mean percent correct and mean response latencies were calculated over each of the stimulus set conditions for each dimension. Response latencies were analyzed for correct responses only.

Insert Table 2 about here

Response Latencies

Table 2 displays the mean response latencies collapsed over subjects for the control, orthogonal, and correlated conditions for the word and voice dimensions for each of the four stimulus variability conditions. A three-way ANOVA was conducted on the latency data for the factors of stimulus dimension, stimulus set, and stimulus variability. A significant main effect of stimulus dimension was obtained $F(1,68) = 13.3$, $p < .001$. Response latencies were faster for classifying the voice dimension than the word dimension. A significant main effect of stimulus set was also obtained $F(2,136) = 178.1$, $p < .001$. Latencies were fastest in the correlated condition, slower in the control condition, and slowest in the orthogonal condition. Newman-Keuls post-hoc tests revealed that performance in the orthogonal condition differed significantly from performance in the control and correlated conditions. A significant interaction of stimulus dimension with stimulus set was obtained $F(2,136) = 15.6$, $p < .001$). Post-hoc tests of this interaction revealed that performance in the orthogonal condition differed as a function of stimulus dimension, while performance in the control and correlated conditions did not. Finally, a significant interaction of stimulus set with stimulus variability

Table 2

Mean response latencies (in msec) collapsed over subjects for all stimulus variability conditions and for word and voice dimensions as a function of stimulus set condition.

Condition	Dimension	Control	Orthogonal	Correlated	Interference
2W x 2T	word	501.7	560.1	478.4	+ 58.4
	voice	470.7	494.2	463.1	+ 23.5
4W x 4T	word	493.2	587.2	482.4	+ 94.0
	voice	484.8	561.8	487.5	+ 77.0
8W x 8T	word	513.9	630.5	466.7	+ 116.6
	voice	473.4	544.6	480.2	+ 71.2
16W x 16T	word	469.5	629.0	444.0	+ 159.5
	voice	460.5	552.5	446.0	+ 92.0

condition was observed $F(6,136) = 6.7$, $p < .001$). Post-hoc tests revealed that performance in the orthogonal condition in the 2W X 2T condition differed significantly from performance in the orthogonal conditions of the 4W X 4T, 8W X 8T, and 16W X 16T conditions, however, no other significant differences between conditions were observed.

These analyses indicate that response latencies varied reliably as a function of the stimulus dimension that was classified and as a function of the stimulus set condition. In order to examine the effects of stimulus set condition on response latencies more closely, a series of one-way ANOVA's was conducted between the control conditions and the orthogonal and correlated conditions for each dimension in all four stimulus variability conditions.

First, we consider the response latencies for the 2W x 2T condition. For the word dimension, the increase in latencies from the control condition to the orthogonal condition was significant $F(1,17) = 8.5$, $p < .01$. This result indicates that when the word dimension was attended to, irrelevant variation in the voice dimension could not be selectively ignored. A significant difference in latencies between the control condition and the correlated condition was not observed. This indicates the absence of a redundancy gain when attending to the word dimension. For the voice dimension, the increase in latencies from the control condition to the orthogonal condition was also significant $F(1,17) = 6.9$, $p < .02$. When the voice dimension is attended to, the irrelevant variation in the word dimension caused interference. Response latencies for the control condition and the correlated condition were not significantly different. Taken together, the presence of orthogonal interference when either dimension was classified is consistent with the hypothesis that both word and voice are processed as integral dimensions.

Next, we consider the response latencies for the 4W x 4T condition. For the word dimension, the increase in latencies from control condition to orthogonal condition was significant $F(1,17) = 53.1$, $p < .0001$. The irrelevant variation in the voice dimension could not be ignored when subjects were required to attend to the word dimension. A significant decrease in latencies from the control condition to the correlated condition was not observed. For the voice dimension, the increase in latencies from the control condition to the orthogonal condition was significant $F(1,17) = 19.6$, $p < .001$. This result indicates that the word dimension could not be ignored when subjects were required to classify the voice dimension. A significant decrease in latencies from control to correlated conditions was also not observed. Thus, for the 4W x 4T condition, orthogonal interference was observed for both dimensions. However, we failed to find significant redundancy gains for either dimension.

For the 8W x 8T condition, the increase in latencies from control condition to orthogonal condition was also significant for both the word dimension $F(1,17) = 55.6$, $p < .0001$ and the voice dimension $F(1,17) = 22.7$, $p < .0001$. A significant decrease in latencies from control condition to correlated condition was observed. but only for the word dimension $F(1,17) = 11.2$, $p < .01$. Thus, while orthogonal interference was present for both dimensions, a redundancy gain was only present when subjects were required to attend to the word dimension. The presence of the significant redundancy gain indicates that voice information was used to facilitate classification of the word on the word dimension while the converse relation was not observed.

For the 16W x 16T condition, a significant increase in latencies from control condition to orthogonal condition was significant for both the word dimension $F(1,17) = 68.5$, $p < .001$ and the voice dimension $F(1,17) = 26.8$, $p <$

.001. And again, as in the previous condition, the decrease in latencies from the control condition to correlated condition was also significant only for the word dimension $F(1,17) = 10.2$, $p < .006$. Thus, as in the 8W x 8T condition, orthogonal interference for both dimensions was observed along with a redundancy gain when subjects attended to the word dimension. However, a complementary redundancy gain was not observed when subjects were required to attend to voice.

Insert Figure 2 about here

Figure 2 shows the amount of orthogonal interference (in msec) for the word and voice dimensions for each of the four stimulus variability conditions. For all four variability conditions, a significant amount of orthogonal interference was obtained when either the word or voice dimension was classified. This result provides evidence for symmetrically integral stimulus dimensions. That is, the degree of interference caused by the irrelevant dimension is about the same for each dimension when stimulus variabilitly is increased. However, a closer examination of the amount of orthogonal interference present for each dimension across all four conditions shows that the amount of interference was greater for the word dimension than for the voice dimension. Thus, perception of the word dimension appears to be subject to more interference by irrelevant variation in the voice dimension than vice-versa. This suggests that while the stimulus dimensions are integral, they do appear to show reliable asymmetry in processing in this task.

Upon further inspection of Figure 2, it also appears that stimulus variability affects performance across conditions. The amount of orthogonal interference obtained for the word and voice dimensions increases as stimulus variability increases. In order to quantify these observations, a two-way ANOVA was carried out to assess the amount of orthogonal interference obtained for the factors of stimulus dimension and stimulus variability condition. A significant main effect of stimulus variability was obtained $F(3,68) = 9.2$, $p < .0001$, indicating that as variability increased the amount of orthogonal interference increased. Post-hoc tests revealed that only the 2W x 2T condition and the 16W x 16T condition differed significantly from one another. A significant main effect of stimulus dimension was also observed $F(1,68) = 12.8$, $p < .001$. Overall, the amount of orthogonal inteference obtained for the word dimension was significantly greater than the amount of interference obtained for the voice dimension. This result supports the asymmetry observed earlier, and suggests that the irrelevant variation in the voice dimension interfered more with processing of the word dimension than vice-versa.

These analyses confirmed both of our earlier observations. The first result was that when variability was increased by increasing the number of words and/or the number of talkers used in the orthogonal set, the amount of orthogonal interference observed became significantly larger. This result demonstrates that stimulus variability either in the voice of the talker or in the acoustic-phonetic information contained in the word-initial consonant of the word affects the time needed to classify both dimensions and that selective attention to one dimension or the other becomes increasingly more difficult as the variability in the dimension increases. The present results show very clearly that the effects of stimulus variability are closely related

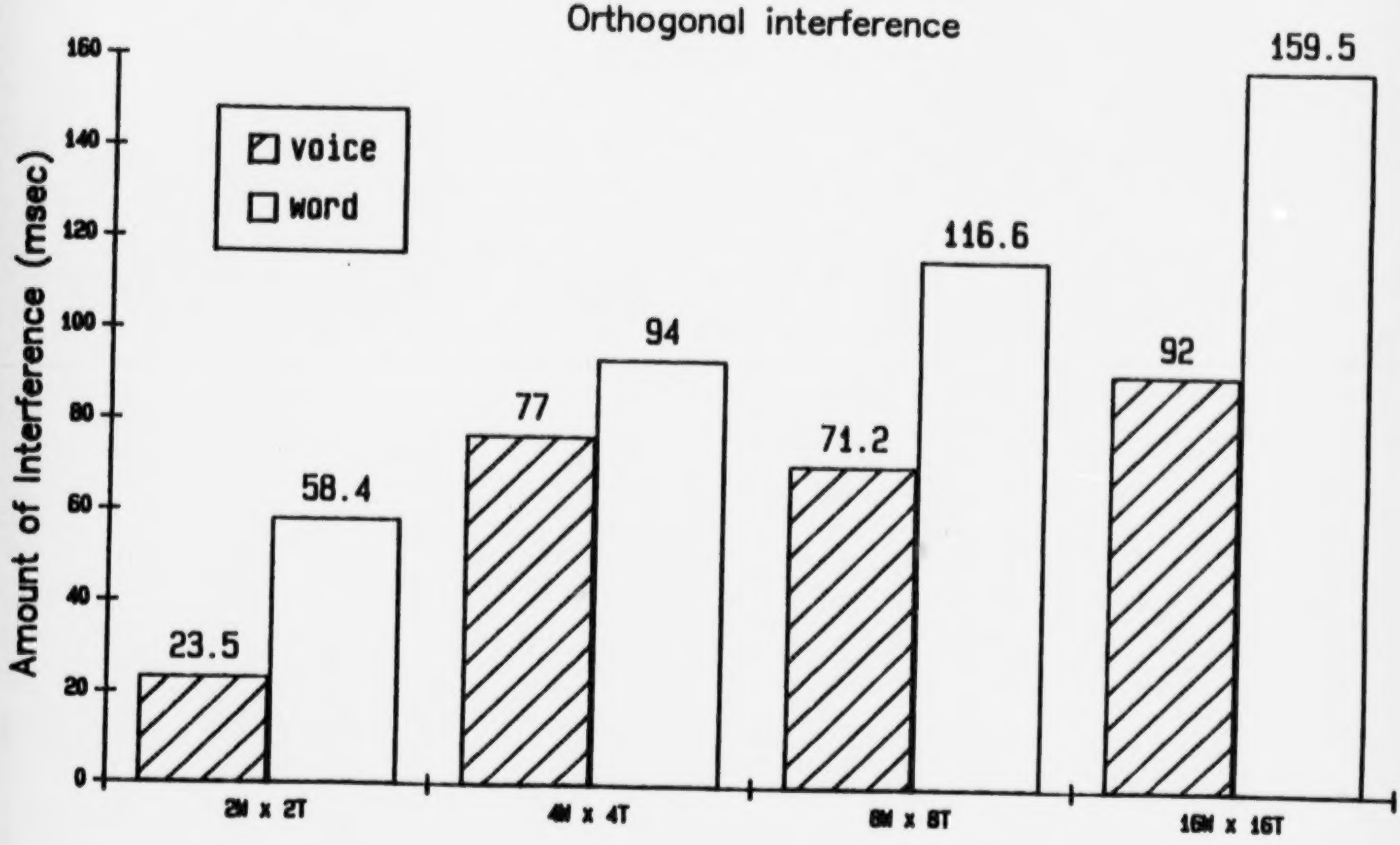


Figure 2. The amount of orthogonal interference (in msec) for all conditions. Interference is shown as a function of word and voice dimensions.

to selective attention to specific stimulus dimensions.

The second result concerns the difference in the amount of interference obtained when comparing performance on the word and voice dimensions. Because the amount of interference was significantly greater for the word dimension across all conditions, the pattern of integrality appears to be asymmetrical in nature. Although selective attention to either dimension was affected by irrelevant variation in the other dimension, variation in the voice dimension interfered with processing of the word dimension to a greater degree than variation in the word dimension affected the processing of voice. This asymmetrical pattern of interference for word and voice dimensions is similar to the asymmetrical pattern found for CV syllables (Wood, 1974).

One explanation of the asymmetrical pattern of interference is related to discriminability of the two dimensions. Under some circumstances, an asymmetrical pattern of interference may be present because of differences in the relative discriminability of the target dimensions (see Eimas et al., 1978; Garner, 1974). If one dimension is inherently more discriminable than the other dimension, the more discriminable dimension may be easier to process when it is relevant but harder to ignore when it is irrelevant. In the present study, the asymmetrical pattern of interference could have been due to the greater discriminability of the voice dimension compared to the word dimension. One method of assessing whether stimulus dimensions in this task differ in discriminability is to compare the response latencies obtained in the control conditions for each dimension. If response latencies are significantly faster in the control condition for one dimension compared to the other, this would support the idea that the faster dimension is more discriminable. Applying this logic to the present results, if the latencies in the voice dimension control condition were faster than those obtained in the word dimension control condition, then the asymmetrical pattern of interference could be explained simply on the basis of discriminability of the individual dimensions.

In order to test this hypothesis, separate one-way ANOVA's were conducted on the latency data for the two control conditions. The results of these analyses indicated that performance for the word and voice dimension control conditions did not differ significantly within any stimulus variability condition. Thus, this result provides support for the claim that the asymmetry we observed was not due to inherent differences in discriminability between the two dimensions but, instead, reflects a real difference in processing between word and voice dimensions.

Insert Table 3 about here

Identification Data Analyses

Table 3 shows the mean percent correct identification data collapsed over subjects for the control, orthogonal, and correlated conditions for word and voice dimensions for all stimulus variability conditions. A three-way ANOVA was conducted on the identification data for the factors of stimulus dimension, stimulus set condition, and stimulus variability. A significant main effect of stimulus set condition was obtained $F(2,136) = 41.1$, $p < .001$.

Table 3

Mean percent correct identification collapsed over subjects for all conditions as a function of stimulus dimension and stimulus set condition.

Condition	Dimension	Control	Orthogonal	Correlated
2W x 2T	word	98.3	97.8	98.9
	voice	99.0	97.2	98.4
4W x 4T	word	98.8	97.2	99.5
	voice	97.7	97.7	99.1
8W x 8T	word	98.2	96.3	98.9
	voice	97.7	96.7	98.2
16W x 16T	word	98.9	97.2	98.9
	voice	98.7	96.8	99.1

Identification was most accurate in the correlated condition, less accurate in the control condition, and least accurate in the orthogonal condition. Post-hoc tests revealed that identification performance in the orthogonal condition differed significantly from performance in both the control and correlated conditions only. No other significant main effects or interactions were obtained.

In considering the identification and the latency data together, the pattern of results suggests that speed-accuracy tradeoffs did not occur in the data. Post-hoc tests showed that identification performance did not differ between the control and correlated conditions while identification performance was worse in the orthogonal condition compared to the other two conditions. Since the increase in latencies from control to orthogonal conditions was not accompanied by an increase in accuracy, and since the decrease in latencies from control to correlated conditions was not accompanied by a decrease in accuracy, further analyses on the data to test for speed-accuracy tradeoffs were not carried out.

The results of the present speeded classification experiment are important in two respects. First, we found that subjects were unable to selectively attend to either word or voice while performing a speeded classification task. When attending to information needed for word classification, the voice information could not be selectively ignored and when attending to voice information, the word-related information could not be selectively ignored. Information concerning word-initial phonetic information and information about the talker's voice appear to be processed together in a mutually dependent, integral manner. Furthermore, the nature of this processing interaction appears to be asymmetrical. The processing of the voice dimension affected phonetic classification more than vice-versa. This processing asymmetry is consistent with the hypothesis that the processing of word-related information is partially contingent on the prior processing of information about voice. That is, although the word and voice dimensions are processed as integral units, the processes extracting information relevant to word recognition may require some information contained in the output of analysis processes operating on classification or encoding of the talker's voice before proceeding. Processing dependencies such as this are consistent with models that operate in a parallel-contingent manner (Turvey, 1973) or in a hybrid serial/parallel manner (Wood, 1974, 1975).

Further evidence for a processing asymmetry between these two dimensions was provided by an examination of the redundancy gains observed in the 8W x 8T and 16W x 16T conditions. In both conditions, a decrease in latencies was observed from the control condition to the correlated condition when subjects attended to the word dimension. In no condition did a significant redundancy gain occur when subjects attended to the voice dimension. This pattern of results is consistent with the idea that voice information is used by the perceptual system in order to classify words. Although significant redundancy gains were not observed for for the 2W x 2T and 4W x 4T conditions, the effects were asymmetrical and were in the predicted direction (see Table 2).

The second important result of the present study concerns the effects of stimulus variability. When stimulus variability was increased, more interference was observed for both word and voice dimensions. The increase in response latencies as a function of stimulus variability is consistent with earlier research showing that variability in the voice of the talker produces detrimental effects on spoken word recognition (Creelman, 1957; Mullennix et al., 1987). Thus, the effects of stimulus variability are not only present in perceptual identification and naming tasks, but apparently also generalize to two-choice speeded classification tasks as well.

One point about the effects of variability that should be mentioned is that, in the present experiment, two sources of variability were manipulated together. It is possible that variability from trial to trial in the acoustic characteristics of the initial consonants may have resulted in greater demands on the perceptual system in encoding phonetic information relevant to the initial consonant. On the other hand, talker variability may have affected performance because of perceptual adjustments related to compensating for the acoustic differences present as a function of changes in talker voice. Since word variability and talker variability were manipulated together, it is difficult to assess whether the increase in orthogonal interference produced by the increase in variability was due to one or both sources of variability. In future experiments, we plan to vary each dimension separately while holding the other one constant in order to dissociate these effects.

Discussion

The results of the present study have several important implications for understanding perceptual normalization in speech perception. Taken together with other recent findings from our laboratory, the present results show that the perceptual processes used to encode voice information are closely related to the processes involved in the encoding of the signal into a phonetic representation. A phonetically-related stimulus dimension and a voice-related stimulus dimension were processed as integral perceptual dimensions. Because neither talker information nor phonetic information can be selectively ignored when attending to specific aspects of a word, we conclude that the processes involved in phonetic coding and the processes involved in perceptual normalization of the talker do not operate independently of one another. This conclusion is supported by the results of Mullennix et al. (1987) who showed that the effects of talker variability interact with degradation of the speech signal. The conclusions of Mullennix et al. that talker normalization processes are intimately related to early perceptual encoding processes in speech is consistent with the present results.

The presence of integrality effects in the speeded classification task also suggests that the processing of voice information is a mandatory encoding operation in speech perception. Because voice information cannot be selectively ignored, selective attention to phonetic information is interfered with by the obligatory processing of voice information. Extending this result to previous research on talker variability, it seems reasonable to suppose that decrements in spoken word recognition incurred by changes in the voice of the talker (Creelman, 1957; Mullennix et al., 1987) may have been due to a failure of selective attention caused by the mandatory processing of talker voice. Whenever the voice of the talker changes, the perceptual adjustments that are made interfere with the allocation of attentional resources to the auditory-to-phonetic coding processes used to encode the phonetic representation. It seems likely that either talker-related processes compete for processing resources that are also used by auditory-to-phonetic coding processes, or else control processes that allocate attentional capacity between the two types of processes utilize additional time or processing resources, when input from different talkers is encountered.

The pattern of integrality effects obtained in the present study provides further insight into the relationship of auditory-to-phonetic coding processes and talker normalization processes. The asymmetric pattern of interference observed, with greater interference caused by the irrelevant variation in the voice dimension, suggests that the analysis of phonetic information contained in word-initial consonants is more dependent on the prior or concurrent analysis of voice information than vice-versa. Asymmetries of this kind have been interpreted in terms of serial and parallel models of processing (see Eimas et al., 1978; Wood, 1974, 1975). In one series of experiments, Eimas et al. (1978) found that the phonetic dimensions of place of articulation and manner of articulation were asymmetrically processed. The processing of place was more dependent on manner than vice versa. This asymmetry was similar to the asymmetry observed in the present study, because a significant amount of interference was obtained for each dimension but it was significantly larger for one dimension than the other. Based on this processing asymmetry, Eimas et al. suggested that the mechanisms of analysis involved in the processing of each phonetic dimension " . . . While functioning in temporally overlapping and interactive fashion, are, to some extent, hierarchially arranged, in that some processes of analysis require the outputs from other analyzers before their own analyses can be completed" (Eimas et al., 1978, p. 18). Hence, Eimas et al. suggested that the phonetic dimensions were processed in what is called a parallel-contingent manner (Turvey, 1973). In Turvey's (1973) model of visual processing, he hypothesized that certain perceptual processes temporally overlap but that one process is contingent on the other. Apparently, because Eimas et al. (1978) found that the processing of both dimensions significantly interfered with one another, and, because place of articulation decisions were more dependent on manner information than vice-versa, it was suggested that the processes extracting each phonetic dimension operate in parallel. Information from the manner of articulation analyzers is used by the place of articulation analyzers in a hierarchially-driven manner.

Wood (1974) also obtained an asymmetric processing relation between two phonetic dimensions. He observed an asymmetry between place of articulation and fundamental frequency, with place of articulation more dependent on fundamental frequency than vice-versa. However, a significant amount of orthogonal interference was observed only for the place dimension. Wood (1974) also obtained significant redundancy gains for both dimensions, a result not observed by Eimas et al. (1978) or in the present results. Wood (1974) argued that the presence of the processing asymmetry and the redundancy gains taken together indicated that a hybrid serial/parallel model of processing was appropriate. The asymmetry in interference suggested that the processing of place of articulation was dependent on pitch information, however, the processing of pitch occurred independently of place. This result is consistent with a serial model of processing in which the processing of pitch is completed before the processing of place information. However, the observed redundancy gains indicated that information from either dimension could be used to assist classification responses, a result that is consistent with a parallel flow of processing for both dimensions. Thus, although the results obtained by Eimas et al. (1978) and Wood (1974) differed, both investigators proposed processing models incorporating serial and parallel components that were in essence very similar to one another in order to account for their findings.

With regard to the present study, the pattern of results we obtained differ slightly from those reported by Eimas et al. (1978) and Wood (1974). The asymmetry in interference we observed resembled the results of Eimas et al. (1978) because we obtained significant interference for both the word and voice dimensions, with the magnitude of interference greater for the word dimension. However, we did observe redundancy gains in some conditions for the word dimension. Because interference was obtained on both dimensions, it is likely that talker normalization processes and auditory-to-phonetic processes operate in parallel. However, because the interference was asymmetric and because the redundancy gains indicated that only the redundant voice information was used to assist classification of the word dimension, it also appears that the auditory-to-phonetic coding processes may be partially contingent on the prior output of the talker normalization processes. Based on our results, we conclude that the processing of talker-related information and phonetic information does not occur in a serial manner. Instead, it appears that processing of these dimensions occurs in a manner best described as parallel-contingent. If there exist multiple information-processing components in speech perception, it is possible that a subprocess operating on encoding the talker's voice and subprocesses operating on phonetically-related auditory information operate in parallel. As these subprocesses are carried out, auditory-to-phonetic processes must wait for at least part of the output from talker-related analysis routines before any further phonetic analysis of the input signal proceeds. However, the talker-related processes use very little information, if any, from the phonetic analyzers. Thus, in effect, a hierarchially-driven contingency of processing exists between talker normalization processes and auditory-to-phonetic coding processes, so that talker normalization processes can be carried out at an earlier functional level in the perceptual system.

With regard to the effects of stimulus variability, the present findings show that an increase in stimulus variability produces increases in response latencies. This result provides converging evidence supporting the results obtained in previous studies on spoken word recognition (Creelman, 1957; Mullennix et al., 1987) and vowel and consonant perception (Assman et al., 1982; Fourcin, 1968; Rand, 1971; Verbrugge et al., 1976; Weenink, 1986) which demonstrated that trial-to-trial changes in the voice of the talker affects speech perception and spoken word recognition. Because an increase in interference was obtained in the speeded classification task by increasing word and talker variability, selective attention to the two target dimensions became more difficult. This decrease in selective attention can be explained in one of two ways. Either talker normalization processes compete for limited-capacity resources also used by auditory-to-phonetic coding processes, or else the operation of control processes that switch control between talker normalization processes and auditory-to-phonetic coding processes is affected, resulting in the need for additional time and/or processing resources. Although our results cannot distinguish between these two alternative accounts, it is clear that the perceptual system compensates in some manner for the acoustic differences due to talker variability and that this compensation produces reliable and robust effects on the processing system.

In summary, the results of the present investigation provide additional information about the relations between talker normalization processes and perceptual processes used to develop segmental phonetic representations. It appears that these perceptual processes are highly interrelated, exhibiting processing dependencies of an asymmetric nature. Selective attention to information in the signal relevant to either type of process appears to be affected by the mandatory processing of the information relevant to the other process. If the pattern of interference observed in the present study had been completely symmetrical, this result would have been consistent with the idea that a single perceptual process produced both sets of results. However, because an asymmetrical processing dependency was observed along with a unidirectional redundancy gain in only one of the dimensions, it is necessary to postulate two separate processes or mechanisms to account for the results. Because of the nature of the processing asymmetry, and, because the redundancy gains indicate that information on the voice dimension can be used to assist word recognition, it appears that these processes are also hierarchially arranged. The analysis of phonetic information is partially contingent on the output of talker voice analyzers. Because a significant amount of interference was observed for both word and voice dimensions, it is reasonable to postulate that the relevant processes overlap temporally and operate in parallel. This description of processing most closely resembles the parallel-contingent model of Turvey (1973), as adopted here to describe the effects of stimulus variability in speech perception. Overall, our results are consistent with the idea that perceptual normalization processes used to encode voice information are intimately related to the early auditory-to-phonetic coding processes involved in speech perception and spoken word recognition.

References

Assman, P.F., Nearey, T.M., & Hogan, J.T. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. Journal of the Acoustical Society of America, 71, 975-989.

Biederman, I.J. & Checkosky, S.F. (1970). Processing redundant information. Journal of Experimental Psychology, 83, 486-490.

Blechner, M.J., Day, R.S., & Cutting, J.E. (1976). Processing two dimensions of nonspeech stimuli: The auditory-phonetic distinction reconsidered. Journal of Experimental Psychology: Human Perception and Performance, 2, 257-266.

Carr, P.B., & Trill, D. (1964). Long-term larynx-excitation spectra. Journal of the Acoustical Society of America, 36, 2033-2040.

Carrell, T.D. (1984). Contributions of fundamental frequency, formant spacing, and glottal waveform to talker identification. Research on speech perception technical report no. 5. Bloomington, IN: Indiana University.

Carrell, T.D., Smith, L.B., & Pisoni, D.B. (1981). Some perceptual dependencies in speeded classification of vowel color and pitch. Perception and Psychophysics, 29, 1-10.

Creelman, C.D. (1957). Case of the unknown talker. Journal of the Acoustical Society of America, 29, 655.

Eimas, P.D., Tartter, V.C., Miller, J.L., & Keuthen, N.J. (1978). Asymmetric dependencies in processing phonetic features. Perception and Psychophysics, 23, 12-20.

Fant, G. (1973). Speech sounds and features. Cambridge, MA: MIT Press.

Felfoldy, G.L. & Garner, W.R. (1971). The effects on speeded classification of implicit and explicit instructions regarding redundant dimensions. Perception and Psychophysics, 9, 289-292.

Fourcin, A.J. (1968). Speech-source interference. IEEE Transactions on Audio and Electroacoustics, ACC-16, 65-67.

Garner, W.R. (1974). The processing of information and structure. Potomac, MD: Erlbaum.

Garner, W.R. & Felfoldy, G.L. (1970). Integrality of stimulus dimensions in various types of information processing. Cognitive Psychology, 1, 225-241.

House, A.S., Williams, C.E., Hecker, M.H.L., & Kryter, K.D. (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. Journal of the Acoustical Society of America, 37, 158-166.

Joos, M.A. (1948). Acoustic phonetics. Language, Suppl. 24, 1-136.

Ladefoged, P. (1980). What are linguistic sounds made of?. Language, 56, 485-502.

Martin, C.S., Mullennix, J.W., Pisoni, D.B., & Summers, W.V. (1987). Effects of talker voice information on recall memory. *Research on Speech Perception Progress Report No. 13*. Bloomington, IN: Indiana University.

Miller, J.L. (1978). Interactions in processing segmental and suprasegmental features of speech. *Perception and Psychophysics*, *24*, 175-180.

Monsen, R.B., & Engebretson, A.M. (1977). Study of variations in the male and female glottal wave. *Journal of the Acoustical Society of America*, *62*, 981-993.

Pastore, R.E., Ahroon, W.A., Puleo, J.S., Crimmins, D.B., Golowner, L., & Berger, R.S. (1976). Processing interaction between two dimensions of nonphonetic auditory signals. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, *267-276*.

Peterson, G.E., & Barney, H.L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*, 175-184.

Rand, T.C. (1971). Vocal tract size normalization in the perception of stop consonants. *Haskins laboratories status reports on speech research*, *SR-25/26*, *141-146*.

Strange, W., Verbrugge, R.R., Shankweiler, D.P., & Edman, T.R. (1976). Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America*, *60*, 213-224.

Summerfield, Q. (1975). Acoustic and phonetic components of the influence of voice changes and identification times for CVC syllables. *Report of speech research in progress*, 2(4). The Queen's University of *Belfast*, *Belfast, Ireland*.

Summerfield, Q., & Haggard, M.P. (1973). Vocal tract normalisation as demonstrated by reaction times. *Report on research in progress in speech perception, No. 2*. The Queen's *University of Belfast, Belfast, Ireland*.

Tomiak, G.R, Mullennix, J.W., & Sawusch, J.R. (1987). Integral processing of phonemes: Evidence for a phonetic mode of perception. *Journal of the Acoustical Society of America*, *81*, 755-764.

Turvey, M.T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, *80*, 1-52.

Verbrugge, R.R., Strange, W., Shankweiler, D.P., & Edman, T.R. (1976). What information enables a listener to map a talker's vowel space?. *Journal of the Acoustical Society of America*, *60*, 198-212.

Weenink, D.J.M. (1986). The identification of vowel stimuli from men, women, and children. *Proceedings 10 from the institute of phonetic sciences of the university of amsterdam*, *41-54*.

Wood, C.C. (1974). Parallel processing of auditory and phonetic information in speech discrimination. *Perception and Psychophysics*, *55*, 501-508.

Wood, C.C. (1975). A normative model for redundancy gains in speeded classification: Application to auditory and phonetic dimensions in speech discrimination. In F. Restle, R.M. Shiffrin, N.J. Castellan, H. Lindman, and D.B. Pisoni (Eds.), *Cognitive theory: Volume 1*. Hillsdale, N.J.: Erlbaum.

Wood, C.C., & Day, R.S. (1975). Failure of selective attention to phonetic segments in consonant-vowel syllables. *Perception and Psychophysics*, *17*, 346-350.

II. SHORT REPORTS AND WORK-IN-PROGRESS

[RESEARCH ON SPEECH PERCEPTION Progress Report No.13 (1987) Indiana University]

Some Observations Concerning English Stress and Phonotactics Using a Computerized Lexicon*

Stuart Davis

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, IN 47405

---

*This research was supported by an NIH Training Grant NS-07134-09 to Indiana University. I wish to thank Mike Dedina and Paul Luce for their help with Lexis on the Symbolics lisp machine and to David Pisoni for his comments.

Abstract

This paper presents some observations concerning English stress and phonotactics that were made with the aid of a computerized lexicon that contains nearly 20,000 entries from Webster's Pocket Dictionary. Observations were made concerning the possible influence of vowel height on stress and on the occurrence of phonotactic restrictions that hold between nonadjacent consonants. By conducting a variety of lexical searches through the 20,000 word lexicon, several types of cases were found in which vowel height had an influence (greater than chance) on whether or not a given syllable received primary stress. One type of case involves stress on nouns like 'minister' and 'semester' where there is an s-cluster between the penultimate and final syllable. Some of these nouns have antepenultimate stress (as in 'minister') while others have penultimate stress (as in 'semester'). It was found that penultimate syllables containing nonhigh vowels were much more likely to receive stress than a penultimate syllable containing a high vowel. Phonotactic constraints were found to occur between nonadjacent consonants in an sCVC sequence. Specifically, it was found that there is a constraint on an sCVC sequence in English words in which the two C's cannot be both labial or both velar. These observations on English stress and phonotactics have not been noted before. These observations, though, should be considered preliminary and should be eventually checked using a larger lexicon.

Some Observations Concerning English Stress and Phonotactics Using a Computerized Lexicon

Introduction

Some current work on English phonology has dealt with questions of stress. Work on stress by Chomsky & Halle (1968), Halle & Keyser (1971), Ross (1972), Liberman & Prince (1977), Hayes (1981), and others have done much in delimiting what the possible stress patterns for English words are. Work on English phonotactics by such researchers as Algeo (1978) and Selkirk (1982) have contributed to our understanding of what the possible sound sequences of English syllables are. Nonetheless, there are many questions about English stress and phonotactics that remain to be considered. In this paper two issues will be explored. The first one, which should be considered more programmatic than definitive, concerns the effect of vowel height on stress; the second one deals with the occurrence of phonotactic constraints holding between nonadjacent consonants and builds on the previous work of Clements & Keyser (1983) and Davis (1984).

The Effect of Vowel Height on the Placement of Primary Stress

In English, several factors are involved in determining the location of primary stress on words. One factor is part of speech. The stress pattern on nouns is different than that of verbs and (unsuffixed) adjectives. For example, verbs and unsuffixed adjectives that end in two consonants normally have primary stress on the final syllable. Examples include the verbs 'avert', 'desert', 'molest', 'usurp' and the adjectives 'adverse', 'covert', 'overt, and 'robust' (where the underlining indicates the stressed syllable). Nouns that end in two consonants normally do not have primary stress on the final syllable. Examples include nouns such as 'concert', 'obelisk', 'object', and 'tempest'. Another factor that is important in determing the location of primary stress is syllable weight. The penultimate syllable of a noun normally receives primary stress if it is heavy; otherwise, if it is light, the antepenultimate syllable receives the stress. (A heavy syllable is a syllable that contains a long or tense vowel, such as /i/, /u/, /o/, /e/, /ay/, /aw/, and /oy/, or ends in a consonant. A light syllable is a syllable that ends in a short or lax vowel). The examples below illustrate this pattern.

(1)	Canada	Arizona	Penobscot
	America	October	synopsis
	labyrinth	horizon	decathlon
	venison	oasis	enigma
	Connecticut	bazooka	babushka
	pyramid	amoeba	eucalyptus
	stamina	hyena	electron

The nouns in the first column all have light penultimate syllables and thus have primary stress on the antepenultimate syllable. The words in the other two columns have heavy penultimate syllables and thus have primary stress on that syllable. The words in the second column have a long or tense vowel in the penultimate syllable and the words in the third column all have a syllable final consonant in the penultimate syllable.

The pattern of primary stress exemplified in (1) accurately captures the location of primary stress on a large number of English nouns. However, there are several classes of nouns that are exceptions to this stress pattern. Three such classes are considered in the first part of this paper. These classes of nouns are those having an (underlying) long vowel or diphthong in the final syllable, those having a penultimate syllable closed by a sonorant consonant (/m/, /n/, /l/, /r/, or the velar nasal), and those having an /s/-plus-consonant cluster immediately following the penultimate vowel. It is argued that in these classes of nouns, vowel height is an additional factor influencing the location of primary stress. This conclusion is based upon examining relevant English words obtained from a computerized lexicon consisting of an edition of Webster's Pocket Dictionary that contains nearly 20,000 words. For each word of the on-line lexicon, there is a phonetic transcription that indicates the location of primary and secondary stress as well as the location of syllable boundaries. Because the computerized lexicon used is somewhat limited in that it only has 20,000 words and does not contain many proper names or place names, the conclusion reached about the effect of vowel height on primary stress must be regarded as only tentative and should be eventually checked using a larger database. The specific findings about the effect of vowel height on the location of primary stress are that a final syllable containing a long high vowel is more likely to receive primary stress than a final syllable with a long mid vowel or diphthong. Furthermore, and somewhat conversely, a high vowel in the penultimate syllable immediately followed by an /s/-plus-consonant-cluster is less likely to receive primary stress than a penultimate syllable containing a nonhigh vowel in the same environment. Finally, a penultimate syllable containing a nonlow vowel immediately followed by a syllable-final sonorant consonant is less likely to receive primary stress than a penultimate syllable containing a low vowel in the same environment. Let us now consider these findings in more detail.

English nouns containing an (underlying) long vowel in the final syllable always have a stress on that syllable. Sometimes this stress is the primary stress on the noun and sometimes it is a secondary stress. This is exemplified by the representative data in (2) in which the nouns on the left all have a primary stress on the final syllable and the nouns on the right all have a secondary stress on the final syllable.

(2)	canteen	centipede
	balloon	costume
	champagne	hurricane
	patrol	cathode
	cologne	chaperone
	July	ally
	demise	decoy
	kowtow	powwow

Phonologists who have tried to analyze English stress patterns have essentially taken one of two strategies in trying to account for which degree of stress surfaces on the final syllables of words like those in (2). One strategy, found in Chomsky & Halle (1968) and Halle & Vergnaud (1987), is to posit a rule that assigns primary stress directly to all final syllables with long vowels and then posit a later rule that has the effect of converting the primary stress into a secondary one, though nouns like those in the lefthand column in (2) would have to be exceptional to this later rule. The second strategy is pursued by Hayes (1981) who would first assign secondary stress to the last syllables of all the nouns in (2) and then would need a special rule (which he does not discuss) to account for the occurrence of primary stress on the final syllables of the words in the lefthand column. In terms of absolute numbers, Hayes's strategy of first assigning secondary stress to the final syllables of the nouns in (2) is superior to the other strategy of first assigning primary stress to them. Specifically, in the on-line Webster's Pocket Dictionary there are approximately 950 nouns (or 70%) that contain a long vowel in the final syllable with secondary stress (excluding compound nouns which are not considered at all in this study), and there are approximately 400 such nouns (or 30%) with primary stress on the final syllable.[1] When the same data are considered based on vowel quality in the final syllable an obvious difference emerges. Nouns containing a long high vowel (/i/ or /u/) in the final syllable receive primary stress more often than nouns with either a long mid vowel or a diphthong in the final syllable. Of the 490 nouns from the on-line lexicon that contain /i/ or /u/ in the final syllable 241 (49%) of them have primary stress on that syllable. Some examples are given in (3).

(3)
bamboo
canoe
fondu
kangaroo
shampoo
caboose
lagoon
monsoon
papoose
taboo

chimpanzee
addressee
antique
jamboree
caffeine
police
ravine
career
fatigue
canteen

Of the remaining 860 nouns with (nonhigh) long vowels in the final syllable only 166 of them (19.3%) have primary stress on that syllable. Specifically, 20% of nouns (121 out of 605) with a long mid vowel in the final syllable have primary stress on that syllable, and 18.3% of nouns (45 out of 245) with a diphthong in the final syllable have primary stress on that syllable. Thus a final syllable with a long high vowel attracts primary stress to a greater degree than final syllables with other long vowels. In order to show that the difference in the stress attracting nature of final syllables with long high vowels is distinct from that of final syllables with other types of long vowels a chi square test was carried out on the data. The test gave significant results ($p < .0001$). A second chi square test was carried out to show that the difference between nouns with long mid vowels in final syllables did not have stress attracting properties significantly different than nouns with diphthongs in the final syllable. The test showed that the difference between these was not significant. Thus, final syllables containing long high vowels are more likely to receive primary stress than final syllables with other types of long vowels.

Other evidence that final syllables with long high vowels are more likely to receive primary stress than final syllables with other types of long vowels comes from the stress properties of monosyllabic suffixes containing long vowels. Productive suffixes, such as -ee or -ese, are much more likely to surface with primary stress than monosyllabic suffixes with other types of long vowels, such as -oid, -ile, and -ite. Representative examples are shown in (4) where the words in the lefthand column all have primary stress on the suffix whereas the words on the righthand column all have secondary stress on the suffix.

(4)
employee
devotee
grantee
journalese

molluscoid
percentile
graphite
metalloid

Thus the evidence from the stress properties of these suffixes further support the contention that, in final syllables, long high vowels behave differently than other long vowels.

Although there is a distinct difference between the stress properties of long high vowels and other long vowels in final syllables, it is far from clear how this difference ought to be incorporated into a formal account of English stress. One possibility is to posit rules that assign primary stress to final syllables with long high vowels and secondary stress to final syllables with nonhigh long vowels. Another possibility is to posit a rule that assigns primary stress to all final syllables with long vowels and then have a later rule that converts the primary stress into a secondary one. Regardless of the exact formal analysis, though, the stress properties of final syllables with high long vowels are significantly different than the stress properties of other final syllables with long vowels.

There are at least two other situations in English where it can be shown that vowel height has an influence on stress placement. One situation involves the occurrence of an /s/-plus-consonant immediately after the penultimate vowel of a word and whether that /s/ makes the penultimate syllable heavy (thus attracting stress onto it). The other situation involves the likelihood of a syllable closed by a sonorant consonant not receiving primary stress because of the height of the vowel in that syllable. For both these situations evidence is presented supporting the contention that (relevant) syllables containing high vowels are less likely to receive primary stress than such syllables with nonhigh vowels. Both these situations are different than the case concerning long vowels in final syllables where it was shown that a syllable with a (long) high vowel is more likely to receive primary stress.

Let us first consider primary stress patterns on words (of at least three syllables) where the penultimate vowel is followed by /s/-plus -consonant. These words vary in whether primary stress falls on the penultimate syllable or on the antepenultimate syllable. Examples in (5a) all display antepenultimate stress while those in (5b) display penultimate stress.[2]

(5) a. armistice
canister
minister
pedestal
Palestine
register
banister
hemistich
orchestra
Philistine
Protestant
talisman

b. apostate
canasta
disaster
hibiscus
intestine
piaster
asbestas
clandestine
fiasco
imposter
Nebraska
semester

The words in (5a) all have antepenultimate stress. This suggests that in these words the /s/ after the penultimate vowel does not close the penultimate syllable but is rather part of the onset of the final syllable. On the other hand, the words in (5b) all have penultimate stress which suggests that the /s/ after the penultimate vowel in these words closes the penultimate syllable and does not form part of the onset of the final syllable. In other words, the forms in (5b) are like those in the third column in (1), where a closed penultimate syllable attracts primary stress. The words in (5a) have the same stress pattern as words like 'algebra', 'discipline', and 'vertebra' in which both members of the consonant cluster after the penultimate vowel syllabify as part of the onset of the final syllable and the penultimate syllable is not heavy.

The data in (5) may lead one to believe that the location of primary stress is random on words where there is an /s/-plus-consonant following the penultimate vowel. Some words are like those in (5a) with antepenultimate stress, and, other words are like those in (5b) with penultimate stress. In fact, of the 46 relevant words that can be included in (5), exactly half (23) have penultimate stress and exactly half have antepenultimate stress. However, when the height of the vowel of the penultimate syllable is considered, a significant generalization emerges. Words with a high penultimate vowel almost always have antepenultimate stress (i.e., they pattern like 5a), and words with a nonhigh penultimate vowel usually have penultimate stress (i.e., they pattern like 5b). Out of the 20 relevant words with a high penultimate vowel 17 have primary stress on the antepenultimate syllable. On the other hand, of the 26 relevant words containing a nonhigh penultimate vowel 20 have primary stress on the penultimate syllable. Thus it seems that in these words vowel height is a factor in determining the location of main stress: A nonhigh vowel in the penultimate syllable is more likely than a high vowel to attract stress when follwed by an /s/-plus-consonant sequence. In order to show that this difference between the stress attracting properties of high vowels and nonhigh vowels for words like those in (5) is statistically significant a chi square test was performed on the data. The test gave significant results ($p < .001$). Hence, it can be concluded that in words like those in (5) vowel height can influence the location of primary stress. However, the rather small number of words that were found with this pattern in the 20,000 word lexicon lessens the significance of this conclusion.

A final situation in which vowel height seems to have an influence on the location of primary stress are cases where heavy syllables closed by a sonorant consonant (and preceded by a single syllable) can fail to receive primary stress although they are in position to receive it. Examples include penultimate closed syllables in nouns and closed antepenultimate syllables followed by the suffix -ary/-ory. Both these types of closed syllables should normally receive primary stress. The examples in (6) show that sometimes they do not receive primary stress. In these words primary stress surfaces on the preceding syllable (' indicates primary stress).

(6)
m'ackintosh
d'avenport
s'epulchre
'ampersand
fr'ankincense
br'igantine
c'avalcade
b'alderdash

l'egendary
d'esultory
d'ysentary
v'oluntary
c'ommentary
'inventory
m'omentary
'adversary

The words in the first column of (6) should all have penultimate stress since the penultimate syllable is closed (Compare with the words in the third column in (1) above). The words in the second column in (6) should have primary stress on the heavy antepenultimate syllable, as is usually the case with words having the suffix -ary/-ory, and which can be illustrated by such words as 'directory' and 'refractory' (in which the stressed syllables are underlined). In order to account for words like those in (6), Kiparsky (1979) and Hayes (1981) propose a rule of sonorant destressing which has the effect of eliminating the primary stress from syllables closed by a sonorant consonant in words of the pattern of (6). What is interesting about the sonorant destressing rule is that often it does not apply, and words of the pattern illustrated in (6) sometimes do surface with primary stress on the syllable closed by a sonorant consonant. Relevant examples of such words are provided in (7).

(7) app'endix
ag'enda
in'ferno
alf'alfa
phal'anges
pen'umbra
comp'ulsory
ad'ultery
rot'unda
ver'anda
am'algam
Nov'ember

The question that emerges from data like that in (6) and (7) is if there is any way of determining which of these words have primary stress on the syllable closed with the sonorant consonant, as in (7), and which of these words do not, as in (6). A search through the online 20,000 word lexicon found 158 relevant words having the pattern illustrated by (6) and (7). [3] Of the 158 relevant words 91 (or 57.6%) of them patterned like (6) in that the syllable closed by the sonorant consonant failed to receive primary stress. The remaining 67 (or 42.4%) did receive primary stress on that syllable like the examples in (7). While these percentages may make it appear arbitrary whether these words have a stress pattern like (6) rather than (7), when vowel height is considered, however, significant differences emerge. In the data there were 35 cases where the relevant syllable closed by a sonorant contained a low vowel. Of these, 25 received primary stress and 10 did not. On the other hand, there were 123 cases where the relevant syllable closed by a sonorant contained a nonlow vowel. Of these, 41 received primary stress and 82 did not. [4] A chi square test was performed to show if the difference between relevant syllables containing a low vowel were different from those containing a nonlow vowel. The findings were significant ($p < .001$). Thus, there is evidence to support the view that relevant syllables closed by a sonorant and containing a low vowel are much more likely to receive primary stress than relevant syllables containing a nonlow vowel. In other words, the rule of sonorant destressing is much more likely to apply if the syllable contains a nonlow vowel rather than a low vowel.

In this section, it has been shown that there are at least three cases where vowel height seems to influence the location of primary stress by greater than chance. However, the exact effect that vowel height has is different depending on the particular case being considered. We have just considered a case where relevant syllables with low vowels tend to receive primary stress. Earlier, we considered a case where penultimate syllables containing a nonhigh vowel immediately followed by an /s/-plus-consonant sequence tended to receive primary stress but penultimate syllables with a high vowel in the same environment tended not to receive stress. We have also considered a case in which a final syllable containing a long high vowel was more likely to receive primary stress than a final syllable with other types of long vowels. It is this last case that is unexpected, because, a priori, one would expect that if vowel quality were to have an influence on stress, syllables containing the more sonorous lower vowels would more likely attract stress than syllables with the less sonorous higher vowels (as in the other two cases). In languages other than English in which vowel height plays a role in stress it is syllables with lower vowels that attract stress, not the higher ones. For example, in the New Guinea language Kobon, described by

Davies (1981), primary stress falls on either the penultimate or final syllable of a verb stem, depending on which has a lower vowel (if both syllables have vowels of the same height, then stress falls on the back vowel). Thus it is surprising that a long high vowel in the final syllable of an English noun is significantly more likely to have primary stress than a final syllable with a long nonhigh vowel. However, perhaps this could be understood under an analysis of English vowels like that proposed by Chomsky & Halle (1968) in which the long high vowels are underlyingly long mid vowels and only become high through a rule of vowel shift that only effects long vowels. Under this view, then, it is not surprising that syllables with these vowels receive stress since they are not underlyingly high (and assuming they only shift to high after stress has been assigned). Note also that in the two other cases in English of vowel height being a factor in stress considered in this paper, syllables with high vowels were not likely to receive stress. Since the high vowels in these two cases are short they would be considered underlyingly high (not mid) in an analysis like that of Chomsky & Halle. Thus, the fact that syllables with long high vowels can be stress attracting whereas syllables with short high vowels never are, seems to provide support for Chomsky & Halle's vowel shift analysis of English, given the assumption that higher less sonorous vowels should not influence stress. These preliminary findings on the relationship between vowel height and primary stress should be considered tentative because the 20,000 word database is limited and does not contain very many proper nouns. Only future work with a computerized lexicon containing a much larger database can verify these initial findings on the effect of vowel height on the location of primary stress.

Cross-Vowel Phonotactic Constraints

In this section, I update previous work (Fudge 1969, Clements & Keyser 1983, and Davis 1984, 1985) that dealt with phonotactic constraints between nonadjacent consonants. These papers pointed to a number of systematic constraints holding between a prevocalic and a postvocalic consonant in English monosyllabic words. One of the strongest of these constraints, and one that has been observed by all three of the above-mentioned researchers, is that there are no monosyllabic words of the form sCVC in which the same noncoronal (labial or velar) consonant flanks both sides of the vowel. Hence, there are no English words like 'spep' or 'skik'. Another constraint, noted by Davis (1984), is that there are no monosyllabic words of the form sNVN (where N can be any nasal consonant). Thus there are no words in English like 'snam' or 'sming'. [5] Here, I point out that these two constraints are in fact more general. Both these constraints are more general in that they are not just constraints on monosyllabic words but they are constraints on any sequence of sCVC (or sNVN) regardless where in the word (or, rather, morpheme) that sCVC sequence (or sNVN sequence) occurs. Also, the constraint on sCVC sequences is not just a constraint on identical consonants flanking both sides of the vowel but on homorganic consonants (i.e., consonants having the same place of articulation) flanking both sides of the vowel.

The constraint on sCVC sequences (in which the C's are identical noncoronal consonants) and sNVN sequences (in which the N is any nasal consonant) is assumed in Davis (1984) as well as in Treiman (1987) to be a constraint on the shape of English syllables. If, in fact, this is a constraint on English syllables one would expect to find words of the form sCVCV (or sNVNV) since the postvocalic C (or N) would not be part of the initial syllable. So, for example one might expect to find words like 'skicky' or 'spapoon' in which the postvocalic consonant is not part of the initial syllable, but not find words like 'skick' or 'spap' in which the postvocalic consonant is part of the initial syllable. If, on the other hand, the constraint on sCVC sequences (and sNVN sequences) is actually a constraint on a sequence of sounds, regardless of whether the sounds are all in the same syllable, then possible words or sequences like 'skicky' or 'spapoon' would be nonoccurring or at least extremely rare. A search was done on the 20,000 word lexicon to see if the sequences sCVC and sNVN occur in any polysyllabic words. The only word in this lexicon in which the sequence sCVC is found (where the C's are noncoronal consonants) is the word 'dyspepsia' where the sequence "spep" occurs. No other such words were found. Polysyllabic words having the sequence sCVC where the two C's are not identical are much more common. A search through the 20,000 word lexicon gives us such words as 'spaghetti', 'scaffold', 'scuba', 'eskimo', and 'episcopal'. Thus it appears that the constraint on sCVC sequences is not really a constraint only holding within a syllable but is a constraint on a sequence of sounds holding within a word.

At first glance, the search through the lexicon of polysyllabic words containing the sequence sNVN suggests that the constraint on the sequence sNVN does not hold for polysyllabic words, unlike the constraint on sCVC sequences. The following twelve words containing the sequence sNVN were found: casement, congressman, dismantle, emplacement, fastening (with the orthographic "e" between the "t" and the "n" being deleted in pronunciation), marksman, placement, pronouncement, replacement, spokesman, statesman, and talisman. However, these words are not monomorphemic; all of these words (with the possible exception of talisman) involve morpheme boundaries between the /s/ and the following nasal consonant. These data thus indicate that the constraint on sNVN sequences (as well as on sCVC sequences) are constraints on a sequence of sounds that hold within morphemes; they can be considered morpheme structure constraints, not word level or syllable structure constraints.

A consequence of the conclusion that these constraints are morpheme structure constraints is that they provide additional evidence against Hooper's (1975) proposal that all morpheme structure constraints are expressible as, and so reducible to, syllable structure constraints. Previously, both Kahn (1976) and Davis (1984) have argued against Hooper's proposal by noting that English has other constraints that are not reducible to syllable structure constraints, such as the prohibition on having two adjacent voiced obstruents monomorphemically. Sequences like /bd/ or /dz/ only occur over a morpheme boundary even though they may be tautosyllabic (as in words like 'nabbed' or 'pods' where the two voiced obstruents in each word are tautosyllabic and a morpheme boundary occurs before the second voiced obstruent). There are no monomorphemic words like these (with the exception of the very low frequency word 'adze'). Thus the constraints on sCVC sequences and sNVN sequences provide additional evidence that English does indeed have morpheme structure constraints.

The constraint disallowing (monomorphemic) sCVC sequences (in which the C's are identical noncoronal consonants) noted by Clements & Keyser (1983) and Davis (1984) actually turns out to be a more general constraint in that the two C's do not have to be identical; rather, they cannot be articulated in the same place in the vocal tract. That is, there are virtually no monomorphemic words in English that have the sequence sCVC where the two C's are either both labial or both velar. The only word in the 20,000 word lexicon that was found to violate this constraint (besides 'dyspepsia') is the word 'skunk'. The words 'skag', 'spam', and 'spumoni' would also violate the constraint although they were not listed in the lexicon. That this constraint really does involve identical place of articulation is made evident when we consider the situation where the two C's in an sCVC sequence are not homorganic. A search through the 20,000 word computerized lexicon revealed that no constraint whatsoever held when the two C's were made at different locations in the vocal tract. For example, there were 58 entries for words having the sequence skVL (where 'k' represents /k/ and L represents a labial consonant) as in 'skip' or 'scuba'; there were 151 words having the sequence skVA (where 'A' is an alveolar) as in 'skit' or 'skate'; and there were 25 words having the sequence skVP (where 'P is a palatal-alveolar consonant) as in 'scotch' or 'sketch'. The fact that there were virtually no words with a velar consonant following an skV sequence is of interest. Moreover, the sequence spV was followed by a velar consonant in 56 entries (eg, 'spike', 'spook'), an alveolar consonant in 196 entries (eg, 'spit', 'speed'), and a palatal-alveolar consonant in 20 entries (eg, 'speech', 'special'); there were virtually no words where a labial consonant followed an spV sequence. Thus it is concluded that the constraint on sCVC sequences originally formulated by Clements & Keyser (1983) and Davis (1984) as a constraint on the occurrence of identical noncoronal consonants is in fact a more general constraint on consonants made in the same place of articulation

Although the constraint against having homorganic (noncoronal) consonants flanking both sides of the vowel in a sCVC sequence seems to be a real constraint of English, it remains somewhat of a mystery why there should be such a constraint. The constraint crucially must include /s/ since there is no constraint on English CVC sequences where the two C's are homorganic. A check through the 20,000 word computerized lexicon found 118 entries for words having (nonnasal) labial consonants flanking both sides of the vowel in a CVC sequence and 138 entries for words having a velar consonant flanking both sides of a vowel in a CVC sequence. Thus this constraint only involves an sCVC sequence and not any CVC sequence. I offer no explanation for why the presence of the /s/ in an sCVC sequence essentially places a restriction on the postvocalic consonant. It is conjectured, though, that while the reason for such a constraint is a mystery, speakers of English make use of them for parsing words in continuous speech. For example, given the constraint on sCVC sequences discussed in this paper, a phonetic sequence like [spaIpleIn] can only be parsed as "spy plane" and not as "spipe lane" nor as a single word. It is quite possible that speakers of English can and do make use of such phonotactic constraints.

In summary, in this paper I have discussed the possible role of vowel height on the placement of stress and the occurrence of phonotactic constraints that hold between nonadjacent consonants. The use of a computerized lexicon allows us to examine these previously unnoticed aspects of English word structure. However, because the 20,000 word lexicon used in this study does have some shortcomings (such as a lack of proper nouns and slang terms) the findings presented in this paper should be considered preliminary. Future work will include trying to verify these findings using a much larger lexicon.

End Notes

1. By long vowel I specifically include the tense vowels /i/, /e/, /u/, and /o/, as well as the diphthongs /ay/, /aw/, and /oy/. I have not included the low vowels because of the uncertainty of whether they should be considered underlyingly tense or not. Perhaps some stressed low vowels in final syllable can be considered as underlyingly long. It is worth pointing out, though, that of the more than 400 nouns that do have a stressed low vowel in the final syllable slightly more than 75% of them have a secondary stress on the final syllable while the remainder have primary stress on the final syllable.

2. It should be noted that certain types of words that have an /s/-plus-consonant sequence after the penultimate vowel have been systematically excluded from this study. These include all verbs since stress patterns on verbs differ from nouns in that normally primary stress on verbs is on either one of the last two syllables. Words with a long vowel in the penultimate syllable have been excluded because such words have penultimate stress, as is illustrated in the middle column of (1). Also, words with suffixes that affect the stress pattern of the whole word have been excluded. These include suffixes like -ic (as in 'parasitic', 'characteristic' or 'sadistic') since words with this suffix virtually always have penultimate stress, as well as suffixes like -scope (as in 'telescope' or 'gyroscope') and -sty (as in 'dynasty' or 'travesty') since these words always have antepenultimate stress.

3. Again certain types of words could not be considered. These include verbs and unsuffixed adjectives which have a different stress pattern than nouns, words with an underlying long vowel in the syllable closed by the sonorant consonant since these syllables would receive stress by virtue of the long vowel, and words containing certain suffixes that have an effect on the stress pattern of the whole word.

4. High and mid vowels are grouped together as nonlow vowels since the difference between them is not significant. Of the 39 cases where the relevant syllable contained a high vowel 16 had primary stress and the other 23 did not. Of the 84 words where the relevant syllable contained a mid vowel exactly one-third of them had primary stress and the other two-thirds did not. Mid vowels included the /r/-colored vowels; sometimes the underlying height of a vowel was determined based on its orthography since the relevant syllable sometimes contained only the reduced form of the vowel. So for example, the word 'voluntary' was considered to have an underlying high vowel in the relevant (the antepenultimate) syllable.

5. Jespersen (1932) noticed the occurrence of a phonotactic constraint in English that holds between the two nonadjacent consonants that flank both sides of a vowel in CVC monosyllables. Specifically, he noted that English has virtually no monosyllabic words of the form gVp (where V stands for any vowel) except for 'gap' and 'gape'. In fact, a check through the 20,000 word online lexicon revealed that the sequence gVp never occurs even in longer words (with the exception of 'guppy' and 'agape'.) Thus there are only four words that have the sequence gVp. Jespersen considered this constraint holding between a prevocalic /g/ and a postvocalic /p/ to be accidental. Jespersen's conclusion about this constraint being accidental is probably correct. A check through the online lexicon revealed that there are no other constraints between a single prevocalic (oral) stop and a following postvocalic stop. For example, there are 28 words with the sequence gVb, 54 words with the sequence bVg, and 34 words with the sequence pVg. It seems,

then, that it is an accidental property of English that there are so few words containing the sequence gVp.

References

Algeo, J. (1978). What consonant clusters are possible? *Word*, *24*, 206-224.

Chomsky, N. & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Clements, N. & Keyser, J. (1983). *CV phonology*. Cambridge: MIT Press.

Davies, J. (1981). *Kobon* (Lingua Descriptive Studies, *3*). Amsterdam: North Holland Publishing Company.

Davis, S. (1984). Some implications of onset-coda constraints for syllable phonology. *Chicago Linguistic Society*, *20*, 46-51.

Davis, S. (1985). *Topics in syllable geometry*. Doctoral dissertation, University of Arizona, Tucson.

Fudge, E. (1969). Syllables. *Journal of linguistics*, *5*, 253-287.

Halle, M. & Keyser, J. (1971). *English stress: Its form, its growth, and its role in verse*. New York: Harper and Row.

Halle, M. & Vergnaud, J.-R. (1987). *An essay on stress*. Cambridge: MIT Press.

Hayes, B. (1981). *A metrical theory of stress rules*. Doctoral dissertation, MIT, Cambridge. (Also distributed by the Indiana University Linguistics Club, Bloomington.)

Hooper, J. (1975). The Archi-segment in natural generative phonology. *Language*, *51*, 536-560.

Jespersen, O. (1933). Monosyllabism in English. *Selected papers in English, French and German by Otto Jespersen* (pp. 384-408). London: George Allen and Unwin, Ltd.

Kahn, D. (1976). *Syllable-based generalizations in English phonology*. Doctoral dissertation, MIT, Cambridge. (Also distributed by the Indiana University Linguistics Club, Bloomington.)

Kiparsky. P. (1979). Metrical structure assignment is cyclic. *Linguistic Inquiry*, *10*, 421-441.

Liberman, M. & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic inquiry*, *8*, 249-336.

Ross, J. (1972). A reanalysis of English word stress. In M. Brame (ed.), *Contributions to generative phonology* (pp. 229-323). Austin: University of Texas Press.

Selkirk, E. (1982). The syllable. In H. van der Hulst and N. Smith (Eds.), *The structure of phonological representations* (Part II, pp. 337-383). Dordrecht: Foris.

Treiman, R. (1987). *Distributional constraints and syllable structure in English*. Unpublished manuscript, Wayne State University, Detroit.

[RESEARCH ON SPEECH PERCEPTION Progress Report No. 13 (1987) Indiana University]

External Validity of Productive Phonological Knowledge: A First Report*

Judith A. Gierut

Speech Research Laboratory
Department of Psychology

Daniel A. Dinnsen and Kathleen Bardovi-Harlig
Department of Linguistics

Indiana University
Bloomington, IN 47405

---

*This research was supported in part by grants from the National Institutes of Health (NS-07134-09 and NS-20976) to Indiana University in Bloomington. We would like to thank the three nonnative speakers for their participation, Jeffrey Harlig for his assistance with data collection, and Steve Chin for data transcription and analyses. Portions of this paper were presented at the 1987 American Speech-Language-Hearing Association Convention, New Orleans.

Abstract

This paper examines the external validity of productive phonological knowledge as a descriptive metric for characterizing the sound systems of adult speakers acquiring a second language. Productive phonological knowledge is a linguistic construct that has been shown to have internal validity for speech disordered children (Gierut, Elbert, & Dinnsen, 1987). In this study, the productive phonological knowledge of three nonnative English speakers was established from independently motivated standard generative descriptions. Results indicated that a given nonnative speaker displayed differential knowledge of target sounds. Also, speakers of the same native language background evidenced differences in knowledge of target sounds. Finally, the "typical" interlanguage phonology of nonnative speakers was comparable to that of speech disordered children in terms of fundamental properties of the sound system; however, differences emerged in the phonological rule account of error productions.

External Validity of Productive Phonological Knowledge: A First Report

Productive phonological knowledge is a linguistic construct that has been recently introduced in the study of children with speech sound disorders (Elbert, Dinnsen, & Weismer, 1984). Descriptively, productive phonological knowledge has been shown to be a factor that may account for individual differences among error patterns of speech disordered children (Dinnsen, 1984; Dinnsen, Elbert, & Weismer, 1980; Gierut, Elbert, & Dinnsen, 1985; Maxwell, 1981). Experimentally, productive phonological knowledge has also been shown to be a factor that may predict sound learning and generalization during treatment (Dinnsen & Elbert, 1984; Dinnsen, Elbert, Weismer, Forrest, & Powell, 1986; Gierut, 1985; Gierut, Elbert, & Dinnsen, 1987). The construct of productive phonological knowledge thus appears to have internal validity for speech disordered children as a metric of both characterization and treatment. Potentially, this construct may also have important pedagogical applications for other language learning populations; to date, however, the extent to which productive phonological knowledge is generalizable to other language learners has not been established. The purpose of this paper is to examine the external validity of productive phonological knowledge as a descriptive tool for characterizing the sound systems of another language learning population, namely, speakers acquiring a second language.

Subjects

Three adult male nonnative speakers of English served as subjects. The speakers resided in the United States for less than 3 months and were enrolled in a semi-intensive English training program at Indiana University. As participants in the semi-intensive program, these speakers displayed sufficient English language proficiency to enroll in University Division courses, but still needed continued language training. The native language of two speakers was Chinese, Wu dialect; the third speaker's native language was Arabic, Gulf dialect. Subjects were selected for study because of poor English pronunciation skills, making them difficult to understand.

Assessing Productive Phonological Knowledge

Each speaker participated individually in two 1-hour sessions during which time spontaneous connected speech and citation form samples were collected. Samples were obtained using an age-appropriate variation of the elicitation procedure developed by Gierut (1985) for children. Sampling procedures allowed a speaker ample opportunity to produce each target English sound in each relevant word position in a minimum of five different exemplars. Procedures also provided an opportunity to produce potential minimal pairs and morphophonemic alternations. Speech samples were tape recorded, phonetically transcribed, and glossed; these then served as the data base for developing standard generative phonological descriptions of each speaker's sound system. Generative descriptions included information about a speaker's phonetic and phonemic inventories, distribution of sounds, use of phonological rules and/or phonotactic constraints, and underlying representation of morphemes. A speaker's underlying representation of morphemes was of most importance in establishing productive phonological knowledge relative to the target sound system.

Results and Discussion

From the generative descriptions, three general findings emerged. First, a given nonnative speaker displayed differential knowledge of target sounds. As an example, one of the native Chinese speakers produced and used [s] appropriately in all relevant contexts; thus, this speaker maintained a target-like underlying representation (i.e., knowledge) of /s/. The same speaker, however, never produced or used [z] in any context as a result of a phonotactic constraint, indicative of a nontarget-like underlying representation of /z/. This speaker also produced errors involving obstruent stops; voiceless and voiced stops were, respectively, aspirated and devoiced. Here, stops were represented underlyingly in a target-like manner, but the application of phonological rules resulted in surface phonetic errors. From this illustration, notice that the phonology of this speaker was described by both phonological rules and phonotactic constraints. The speaker's phonological knowledge relative to the target was characterized by target-like knowledge (in the case of /s/), target-like knowledge affected by phonological rules (in the case of the stop series), and nontarget-like knowledge (in the case of /z/).

Second, across speakers of the same native language, differences in phonological knowledge were observed. For instance, both native speakers of Chinese exhibited errors involving target /r/. One of the speakers maintained a target-like underlying representation altered by a phonological rule of word-final deletion as supported by morphophonemic alternations between [r] and null. The other speaker maintained a nontarget-like underlying representation as a result of a phonotactic constraint; [r] was absent from the phonemic inventory. Notice that, although the speakers shared the same language background, their phonological knowledge of the target phonology differed in terms of both the nature of the underlying representation and the phonological rule account of errors.

Thus, within and across nonnative speakers of English, differences in productive phonological knowledge emerged. These differences in phonological knowledge could not be accounted for by contemporary principles such as the Contrastive Analysis Hypothesis (Lado, 1957) or the Markedness Differential Hypothesis (Eckman, 1977, 1985). These principles aim to predict accuracies and omissions in the phonologies of nonnative speakers relative to the target language by using, for example, the nature of the native language or typological markedness as a guide. Each speaker therefore had a unique "interlanguage" phonology (Selinker, 1972), distinct from both the native and the target language phonologies. Parallel findings have also been reported for phonologically disordered children (Camarata & Gandour, 1984; Dinnsen et al., 1980; Gierut, 1985, 1986a; Gierut & Elbert, 1983; Gierut et al., 1985; Maxwell, 1981). Children likewise exhibited differences in phonological knowledge and maintained sound systems independent of the target (adult) system. From descriptions of this type, individual differences can be identified based on the nature of speakers' productive phonological knowledge.

A third observation related to comparisons of the nature of a "typical" or average interlanguage phonology to that of an average speech disordered phonology. Figure 1 presents the average interlanguage phonology of the three speakers of this study compared to an average speech disordered phonology as reported by Dinnsen (1986). Notice that approximately half of the typical interlanguage phonology was completely ambient-like (i.e., target appropriate); the other half was associated with errors in sound production and use. The greatest portion of errors was associated with phonological rules (i.e., 41% of the interlanguage phonology). In contrast, only a little more than one-third of an average speech disordered phonology was completely ambient-like. Here, errors in production were attributed primarily to phonotactic constraints, rather than phonological rules. In particular, inventory constraints accounted for the greatest portion of errors (i.e., 30% of a disordered phonology). In terms of underlying representations, only 13% of the interlanguage phonology was characterized as nonambient-like in comparison to 42% of a speech disordered phonology. For the most part, second language learners maintained underlying representations that were comparable to those of native language speakers; whereas, speech disordered children maintained underlying representations that were different from those of adults. Observed differences in the nature of typical sound systems of these two populations may be related in part to differences in sample size. The average interlanguage phonology was calculated on data from 3 nonnative subjects; the average speech disordered phonology was based on data from 29 subjects. The literature on interlanguage phonological systems, however, supports these general findings; namely, the application of phonological rules outweighs that of phonotactic constraints in accounting for the production errors of nonnative speakers (Dickerson, 1975; Eckman, 1981a, 1981b; Hammerly, 1982).

Insert Figure 1 about here

Implications for Future Research

From this report, it appears that the construct of productive phonological knowledge may be a valuable descriptive metric for identifying individual differences among sound systems of nonnative speakers and for characterizing the fundamental properties of interlanguage sound systems. It remains to be determined, however, whether productive phonological knowledge will also be an important factor in predicting learning following second language instruction. In light of research on phonologically disordered children (Dinnsen & Elbert, 1984; Gierut et al., 1987), two specific hypotheses associating phonological knowledge and learning in second language learners must be evaluated. These are: (a) a nonnative speaker's performance on sounds of which s/he has most knowledge (e.g., sounds affected by phonological rules) will be better than performance on sounds of which s/he has least knowledge (e.g., sounds restricted by inventory constraints) and (b) second language instruction on most knowledge (e.g., elimination of phonological rules) will result in limited restructuring of the overall sound system, whereas instruction on least knowledge (e.g., elimination of inventory constraints) will result in widespread restructuring. Preliminary studies addressing the first hypothesis have been reported in the second language literature (Briere, 1966; Hammerly, 1982); however, these studies were descriptive in nature and presented conflicting results. We are currently involved in experimental research to evaluate both of these hypotheses. Continued descriptive and experimental research of this type will more firmly establish the external validity of the construct of productive phonological knowledge.

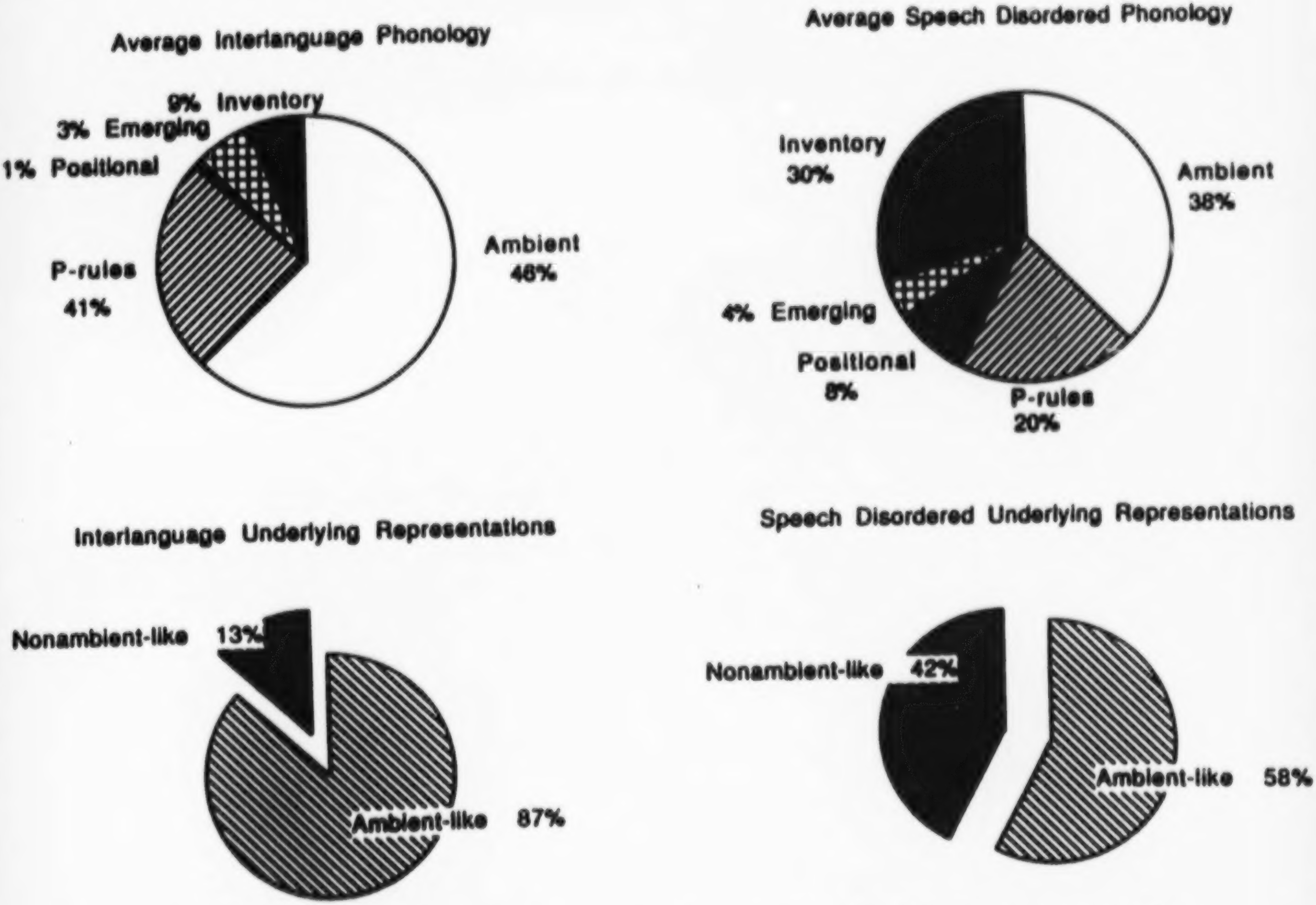


Figure 1. Average interlanguage phonology compared with average speech disordered phonology. Average sound systems were calculated using procedures described by Dinnsen (1986).

References

Briere, E.J. (1966). An investigation of phonological interference. Language, 42,, 768-796.

Camarata, S., & Gandour, J. (1984). On describing idiosyncratic phonologic systems. Journal of Speech and Hearing Disorders, 49, 262-269.

Dickerson, L.J. (1975). The learner's interlanguage as a system of variable rules. TESOL Quarterly, 9, 401-407.

Dinnsen, D.A. (1984). Methods and empirical issues in analyzing functional misarticulations. In M. Elbert, D.A. Dinnsen, and G. Weismer (Eds.), Phonological theory and the misarticulating child (ASHA Monographs 22, pp. 5-17). Rockville, MD: ASHA.

Dinnsen, D.A. (1986, November). Fundamental characteristics of disordered phonological systems. Presented to the American Speech-Language-Hearing Association, Detroit.

Dinnsen, D.A., & Elbert, M. (1984). On the relationship between phonology and learning. In M. Elbert, D.A. Dinnsen, and G. Weismer (Eds.), Phonological theory and the misarticulating child (ASHA Monographs 22, pp. 59-68). Rockville, MD: ASHA.

Dinnsen, D.A., Elbert, M., & Weismer, G. (1980). Some typological properties of functional misarticulation systems. In W.O. Dressler (Ed.), Phonologica 1980 (pp. 83-88). Innsbruck: Innsbrucker Beitrage Zur Sprachwissenschaft.

Dinnsen, D.A., Elbert, M., Weismer, G., Forrest, K., & Powell, T.W. (1986, November). Project report on phonological knowledge and learning patterns. Presented to the American Speech-Language-Hearing Association, Detroit.

Eckman, F. (1981a). On predicting phonological difficulty in second language acquisition. Studies in Second Language Acquisition, 4, 18-30.

Eckman, F. (1981b). On the naturalness of interlanguage phonological rules. Language Learning, 31, 195-216.

Elbert, M., Dinnsen, D.A., & Weismer, G. (Eds.) (1984). Phonological theory and the misarticulating child (ASHA Monographs 22). Rockville, MD: ASHA.

Elbert, M., & Gierut, J.A. (1986). Handbook of clinical phonology: Approaches to assessment and treatment. San Diego: College-Hill Press.

Gierut, J.A. (1985). On the relationship between phonological knowledge and generalization learning in misarticulating children. Doctoral dissertation, Indiana University, Bloomington. (Distributed by the Indiana University Linguistics Club, 1986).

Gierut, J.A. (1986a, November). On determining children's lexical representations. Presented to the American Speech-Language-Hearing Association, Detroit.

Gierut, J.A. (1986b). On the assessment of productive phonological knowledge. *NSSLHA Journal*, *14*, 83-101.

Gierut, J.A., & Elbert, M. (1983, November). Phonological knowledge and the selection of training targets. Presented to the American Speech-Language-Hearing Association, Cincinnati.

Gierut, J.A., Elbert, M., & Dinnsen, D.A. (1985). On characterizing phonological knowledge in disordered sound systems. *Research on speech perception progress report no. 11*. Bloomington, IN: Speech Research Laboratory, Department of Psychology, Indiana University.

Gierut, J.A., Elbert, M., & Dinnsen, D.A. (1987). A functional analysis of phonological knowledge and generalization learning in misarticulating children. *Journal of Speech and Hearing Research*, *30*, 462-479.

Hammerly, H. (1982). Contrastive phonology and error analysis. *International Review of Applied Linguistics*, *20*, 17-32.

Maxwell, E.M. (1981). *A study of misarticulation from a linguistic perspective*. Doctoral dissertation, Indiana University, Bloomington, IN. (Distributed by the Indiana University Linguistics Club, 1982).

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, *10*, 209-231.

Effects of Changes in Spectral Slope on the Intelligibility of Speech in Noise*

Robert I. Pedlow

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, IN 47405

---

*This research was supported, in part, by the Armstrong Aerospace Medical Research Laboratory Contract No. AF-F-33615-86-C-0549 to Indiana University in Bloomington. I thank Professor David Pisoni for originally suggesting the problem to me and his help and comments throughout this investigation.

Abstract

Previous studies have shown that attenuation of the first formant results in enhanced speech intelligibility in noise. The objective of the present study was to distinguish between two possible explanations of this finding. One explanation suggests that it is caused by upward spread of masking in normal speech perception. The other explanation suggests that attenuating F1 increases the information bearing elements relative to the total speech energy in the signal. Stimulus materials were a set of 90 phonetically balanced (PB) words produced by one male speaker. Two linear phase filters were designed. The highpass filter had a pass band from 1100 to 5000 Hz and a stop band from 0 to 850 Hz. The low pass filter used the same bands but reversed the stop and pass bands. The attenuation levels of the stop band were 10, 20, 40 and 80 dB relative to the unfiltered signal. Items were equated for RMS energy after filtering. Processed stimuli were presented to subjects in a perceptual identification paradigm at 0 dB S/N. Articulation index values were also computed for the test items. Fitting transfer functions to the data showed that the improvement in intelligibility obtained by high pass filtering could not be explained from the increase in articulation index values alone. This finding supports the hypothesis that attenuating the F1 in speech results in an increase in intelligibility due to a release from upward spread of masking in the speech signal.

Effects of Changes in Spectral Slope on the Intelligibility of Speech in Noise

It has long been known that a variety of changes occur in the acoustic characteristics of speech when speakers are required to talk in a noisy environment (Lane & Tranel, 1971; Webster & Klump, 1962; Kryter 1946). The present investigation was designed to explore the perceptual consequences of one of these changes, the tendency for the high frequency components of the signal to be relatively emphasized compared to the low frequency components (Webster & Klump, 1962).

From the literature it seems plausible that the acoustic changes observed in speech produced in noise are a consequence of one possible strategy adopted by speakers who are attempting to optimize the intelligibility of their speech (Lane & Tranel, 1971). Indeed, Dreher and O'Neill (1957) found that speech originally produced in noise was more intelligible when presented to listeners in noise than speech produced in quiet. This finding has been replicated in our laboratory recently (Summers, Pisoni, Bernacki, Pedlow, & Stokes, in press). It is thus of some interest to consider the perceptual consequences of the changes in the speech signal that occur when a talker is speaking in noise. The effect which this study focuses on is the relationship of spectral tilt to intelligibility.

There is a large body of research in speech perception and in communications engineering that examines the use of high pass filtering as a method of enhancing the intelligibility of speech in noise. In communications engineering, a substantial effort has been devoted towards development of speech signal enhancement techniques using highpass filtering. Studies by Thomas and Ravindram (1971) and Thomas and Niederjohn (1968,1970) have shown that high pass filtering produces enhanced intelligibility of speech in noise. Thus, there is reason to believe that changes in the relative distribution of energy in the speech spectrum may have some consequences for intelligibility. The problem which the present study focuses on is understanding the perceptual basis of this relationship.

Two approaches have been taken in the literature to understanding the relationship between intelligibility and distribution of speech energy in the frequency spectrum. The first developed from research examining the contribution of different frequency bands to speech intelligibility (Licklider & Miller, 1951). From empirical measurements, a formula known as the articulation index or A.I. was developed using the following model. The signal is first divided into 20 frequency bands which contribute equally to intelligibility (French & Steinberg, 1947). For each band, a S/N ratio is specified based on the peak signal and noise intensity values for the band. The A.I. is then computed as a weighted sum of these ratios. The resulting index provides one method of predicting observed intelligibility scores. This formula assumes that the contribution of each frequency band to the obtained intelligibility is independent of the other bands. Furthermore, it assumes that the information needed for speech intelligibility is equivalent across frequencies.

The other approach in the literature examines the problem in terms of frequency masking. It is a very well known finding in the psychoacoustic literature that for the detection of simple tones, high energy, low frequency tones may function as maskers of high frequency tones. This phenomenon is often referred to as "upward spread of masking " (see, for example, Wegel & Lane, 1924). In normal speech, the amplitude of the average long term speech spectrum decreases by approximately 6 dB per octave increase in frequency. There is some reason then to suppose that in speech perception, the low frequencies may mask the high frequency components. Except for a study by Nye, Nearey, and Rand (1974), relatively little work has been reported in the speech perception literature which bears on this question. There is some research in the literature which demonstrates the existence of masking in speech perception. However, these studies are principally concerned with phase effects rather than frequency domain effects (Hirsh, 1948). A study by Flanagan and Saslow (1958) examined pitch discrimination for synthetic vowels and found that the difference limen increased slightly with increased intensity, in contrast to pure tone difference limens. The authors speculated that this could be a consequence of the low frequency components in the vowel sounds functioning to mask the high frequency components. Two more recent studies by Rand (1973) and Nye et al. (1974) examined dichotic release from masking in the perception of synthetic speech stimuli. The basic conclusion from both of these studies was that there is a significant effect of masking of the high frequency components by the low frequency components in speech. The results showed dichotic release from masking on the order of about 20dB.

The objective of the present study was to distinguish between two hypotheses that have been proposed to explain the observed enhancement produced by high pass filtering. One hypothesis is that:

> "a speech signal which has been passed through a highpass filter which at least to some degree deemphasizes the low frequency content of speech will contain more "intelligibility information" per unit of speech energy than the original signal."
>
> (Thomas & Ohley, 1972)

Thus speech that has been high-pass filtered will be more intelligible in a given noise background than normal speech. The other hypothesis explains the observed effect in terms of the upward spread of masking, and suggests that attenuation of the first formant compensates, to some extent, for masking of the higher formants that is normally present.

The present experiment employed filtering techniques to emphasize either a high band from 1100 Hz to 5 khz or a low band from 0 to 850 Hz. A baseline condition with no filtering was also included. The band edge frequencies used were similar to those used by Rand (1973) and Nye et al. (1974). For a male talker, they divided the speech signal between the F1 and F2 formants. The resulting stimulus materials were then presented in broadband white noise. We hypothesized that the high-pass condition would result in an increase in intelligibility relative to the baseline, whereas the low pass condition would result in a decrease in intelligibility. Articulation index values or (AI) were also computed for the stimulus materials in order to assess whether the performance in all conditions would be equally well predicted by the AI. If there is a release from upward spread of masking, several of the high pass conditions should be poorly predicted by the AI.

Method

Subjects. Five subjects from the Speech Research Laboratory paid subject pool took part in the listening test. These subjects were all native speakers of English with no reported speech or hearing disorder at the time of testing. Subjects were paid for their services at an hourly rate.

Design. The experimental variables were (1) filter type, (2) level of stop band attenuation, and (3) days of training. The basic design was a 3x4x5, with three filter conditions, i.e., high pass, low pass and a baseline with no filtering, four levels of attenuation, (10, 20 40 and 80 dB stop band attenuation) and five days training. Within blocks, the stimulus materials were randomized for filter type and band attenuation level.

Materials. The stimulus materials consisted of a set of 90 phonetically balanced words (PB words) produced by one male speaker. These items were a subset from a larger database of PB words used as testing materials in the Speech Research Laboratory. The words were processed through one of two different digital filters which attenuated either the high frequency band (1100 Hz to 5 khz) or the low frequency band (0 Hz to 850 Hz). The levels of band attenuation used were (10, 20, 40, and 80 dB).

The filters were linear phase filters which removed any possible confounds due to effects of phase in masking. The filter shapes were designed using the ILS system, a standard digital signal processing software package. This was done by selecting a set of coefficients for the filter equations and then examining visually the resulting impulse response plot. The coefficients were then adjusted to give the closest match to the desired frequency response.

Insert Figures 1 and 2 about here

The stimulus set thus consisted of 90 English words that were divided equally among (four band attenuation levels) x (two filter conditions, high and low pass) + (a baseline no filter group). After the stimulus materials were filtered, they were processed again using a second program that matched the stimulus materials for overall RMS amplitude. Figures 1 and 2 show the resulting long-term average spectrum for the items in high and low pass condition, (- 20 dB stopband attenuation). Because the items were equated for overall RMS energy, the resulting long term spectra were basically symmetric about the cut off point of the stop band. The words were presented against a white noise background at 0 db S/N. The signal levels of speech and noise were set at 70 dB SPL.

Procedure. The experiment was controlled by an on-line program implemented on a PDP-11/34 computer. In the basic experimental task, subjects heard each word binaurally through matched and calibrated TDH-39 headphones and typed in their responses on a computer terminal. After each trial, the correct word appeared on the CRT screen in place of the response entered by the subject. Subjects were run together in small groups in a sound-treated room used in listening experiments. Subjects were trained on the task over a period of five days.

Articulation Index Measurement

Measurement of the articulation index (AI) values of the stimuli used the approach developed by French and Steinberg (1947). The first stage in calculating the articulation index scores requires computation of the mean signal intensity in each of the twenty equal articulation bands. There was one approximation required in the present study. Because the original stimuli

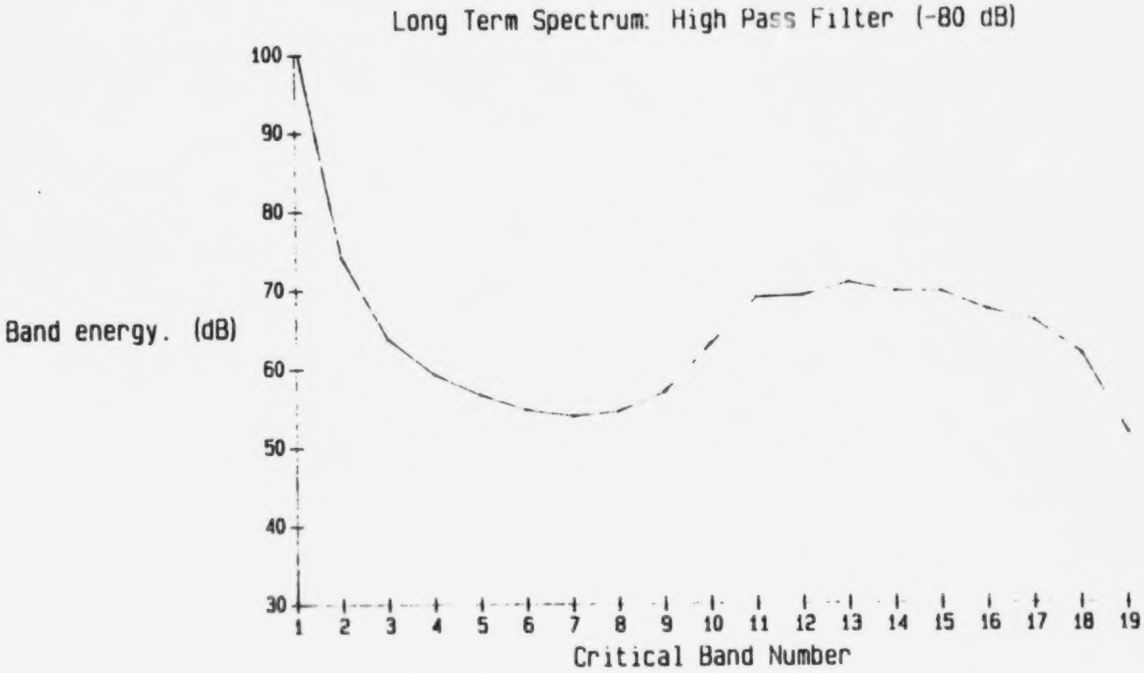


Figure 1. Long term spectrum for Condition 4: High Pass filtered with 80 dB. of stopband attenuation.

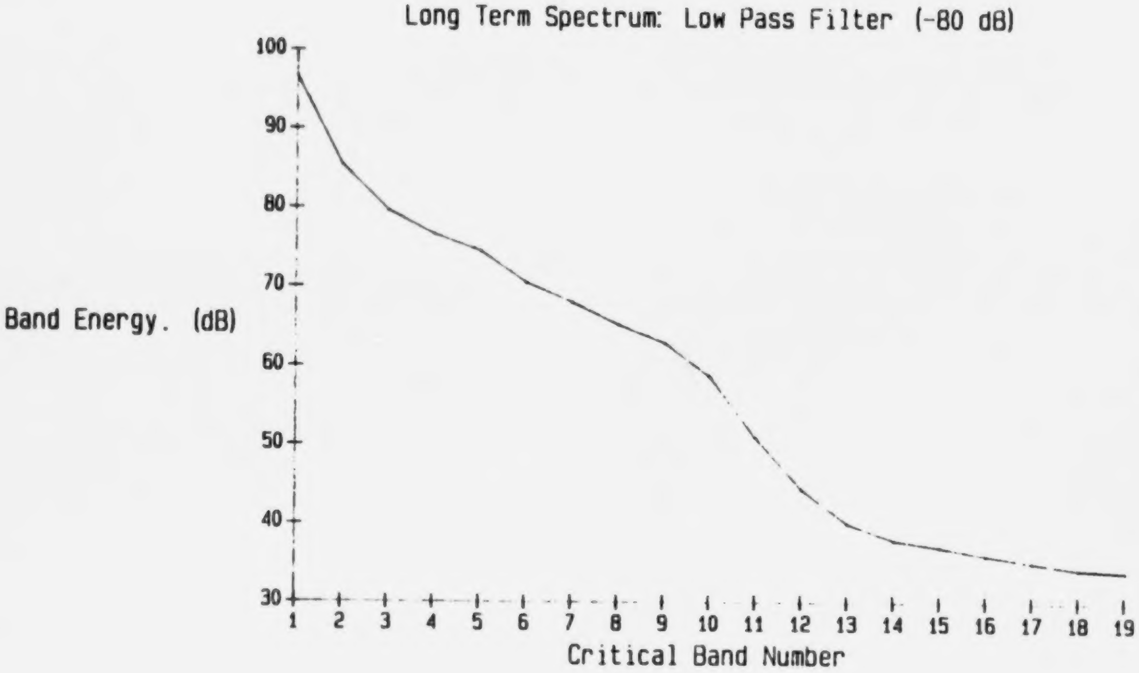


Figure 2. Long term spectrum for Condition 8: Low Pass filtered with 80 dB. of stopband attenuation.

were sampled at 10 kHz, the bandwidth available was 5 kHz which only includes 19 of the twenty bands specified by French and Steinberg. The stimulus files used in the identification task were analyzed using the ILS API program. This program implements LPC analyses of speech waveforms (Markel and Gray, 1976). The LPC parameter setting used in the analysis were (No pre emphasis: Hamming window: 14 coefficients in the LPC equations: 25.6 ms window: 12.8 ms step size).. These analysis files were then analyzed using a second program that computed 128 point FFT's on each LPC analysis vector. From these data, the program computed amplitudes in dB in each of the critical bands. A sample of the output from the white noise source was digitized and analyzed using the same methods to measure the noise level in each band. These measurements showed that the noise source was spectrally flat over the frequency range used. The formula used to compute the articulation index was as follows:

$$AI = \mathrm{sum}(i=1,19)\ Wi*((SNR+12)/30)$$

AI is the articulation index; SNR is the signal to noise ratio in each of the first 19 critical bands; and Wi are the French and Steinberg values for the importance weightings of the critical bands.

Results

The major expectations for the perceptual data were that subjects would show enhanced performance relative to the control condition for some of the high pass conditions and reduced performance relative to control for the low pass conditions. The major results are presented in Figure 3 which shows the mean percent correct values by condition and level collapsed over days of training. Each bar in the figure reflects the mean performance on ten items, by three repetitions by five days of training.

Insert Figure 3 about here

The data in Figure 3 show a significant advantage for the first two highpass filtered conditions, (-10 and -20 dB.) relative to the control condition $F(1,14) = 113.29$, $p < .001$, $F(1,14) = 80.57$, $p < .001$. The other two highpass conditions were not significantly different from the control.

One of our initial concerns in designing this study was that the enhancement effect produced by high pass filtering might be confounded with learning effects on the task. To control for this problem, subjects were trained on the task over five days with three repetitions of the test stimuli on each day of training.

The results shown in Figure 4 display performance across the three major conditions, i.e., control, highpass filtered, and lowpass filtered, against days of training. For the two filtered conditions, each data point reflects the mean performance for three repetitions of twenty items, i.e. collapsed across band attenuation level, for the 10 and 20 dB conditions, for each day of training. For the control condition, each data point reflects performance on ten items by three repetitions for each day of training.

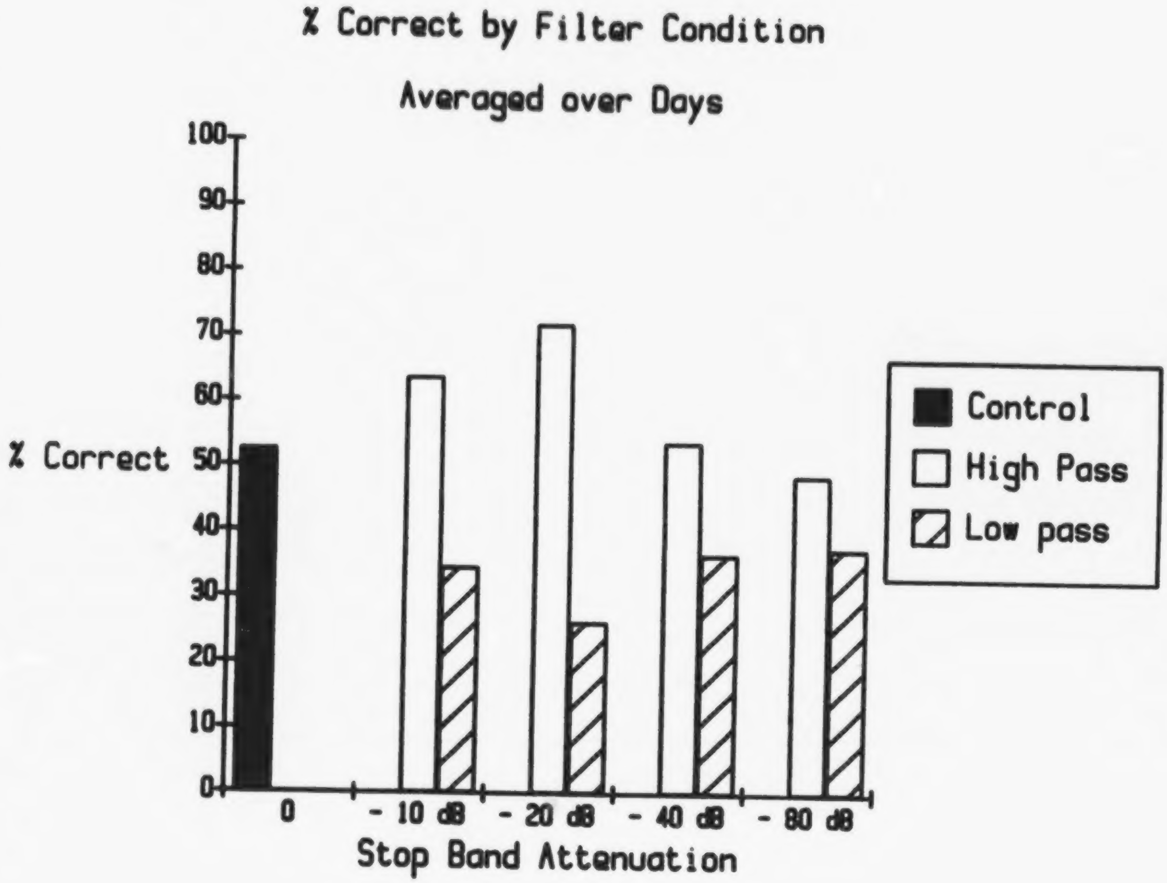


Figure 3. Percent correct identification by filter condition and stop band attenuation level.

Insert Figure 4 about here

Because subjects heard the same items each day and received feedback about the accuracy of their responses, there is a steady increase in percent correct over the first three to four days of training. Performance appears to be approaching an asymptotic level around day four or five. These training data are of interest in several respects. However, for the purpose of the present study the point to note is that the difference between the conditions is consistent over training days.

The method used to compare the perceptual data to the articulation index predictions involved fitting a transfer function relating the articulation index measurements of the stimuli to the obtained perceptual identification data. The transfer function used was the version developed for words rather than nonsense syllables proposed by Humes, Dirks, Bell, Ahlstrom, and Kincaid (1986).

S**(1/n) = (1 - 10**(-a/q))

Where a is the articulation index, S is the sound score scaled as a percent correct between zero and one, n corresponds to the number of sound units, and q is an arbitrary fitting constant. The values for q and n were obtained through simple least squares minimization of the function above to the observed data points.

The results presented in Figure 5 show the observed performance against the AI predictions. The two dotted lines show the plus and minus 2SD curves. The control condition was not used in the curve-fitting calculations.

Insert Figure 5 about here

The x axis in each case is the articulation index, which has a possible range between zero and one. The y axis is the percent correct expressed as a decimal value. Overall, the articulation index appears to predict reasonably well the observed performance data. As expected, the low pass conditions show lower articulation index scores and lower performance results than the high pass conditions. One would expect to find an effect of release from masking from attenuation of the F1 in some of the high pass filtered conditions. From the figures, high pass filtered speech with 20 dB stop band attenuation is consistently the worst predicted by the AI of all the experimental conditions. For four out of the five days of training, this condition is either at or beyond the (+2 SD) curve. Thus for this condition, the speech shows an improvement in intelligibility greater than the AI predicts. In considering these results, it is relevant to note that under 0 dB S/N conditions the variability in subjects performance is relatively high. Thus in more favourable S/N levels one would expect the apparent advantage to be somewhat more marked than the results shown in these figures.

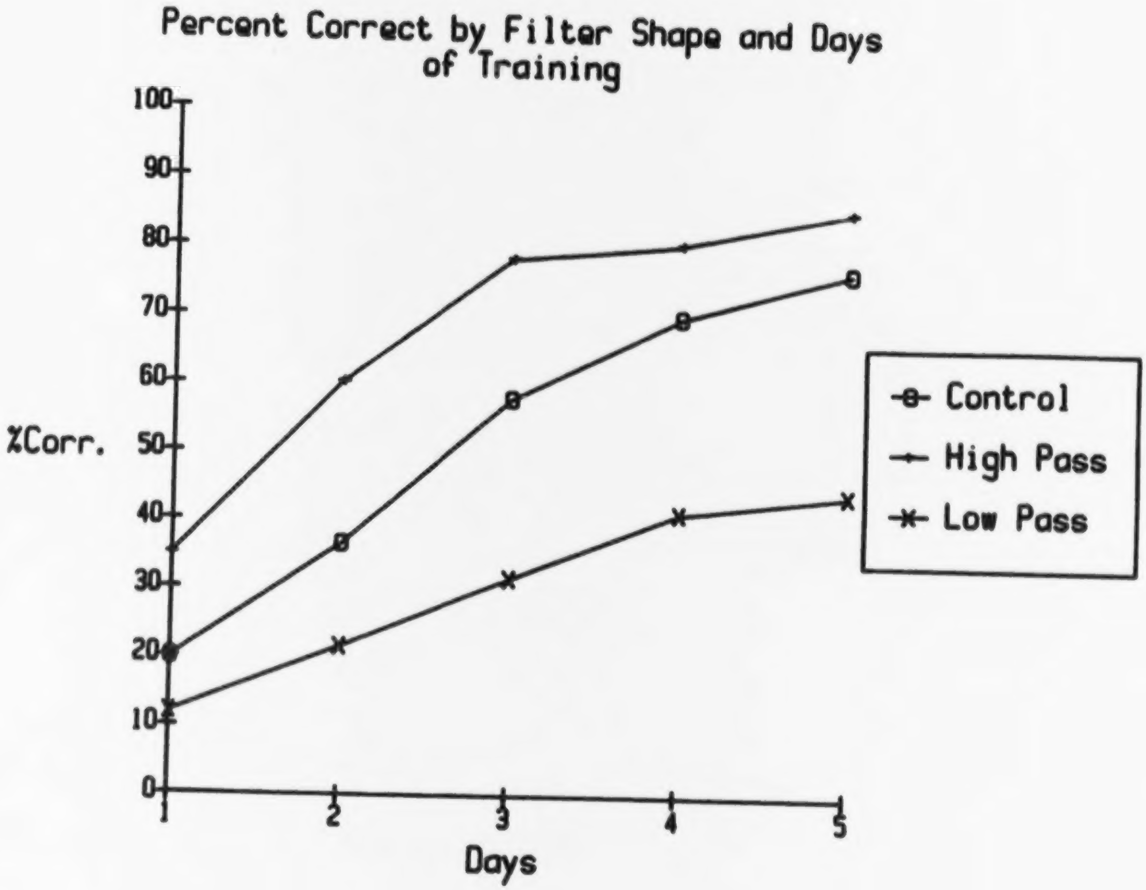


Figure 4. Percent correct identification by filter condition over days of training.

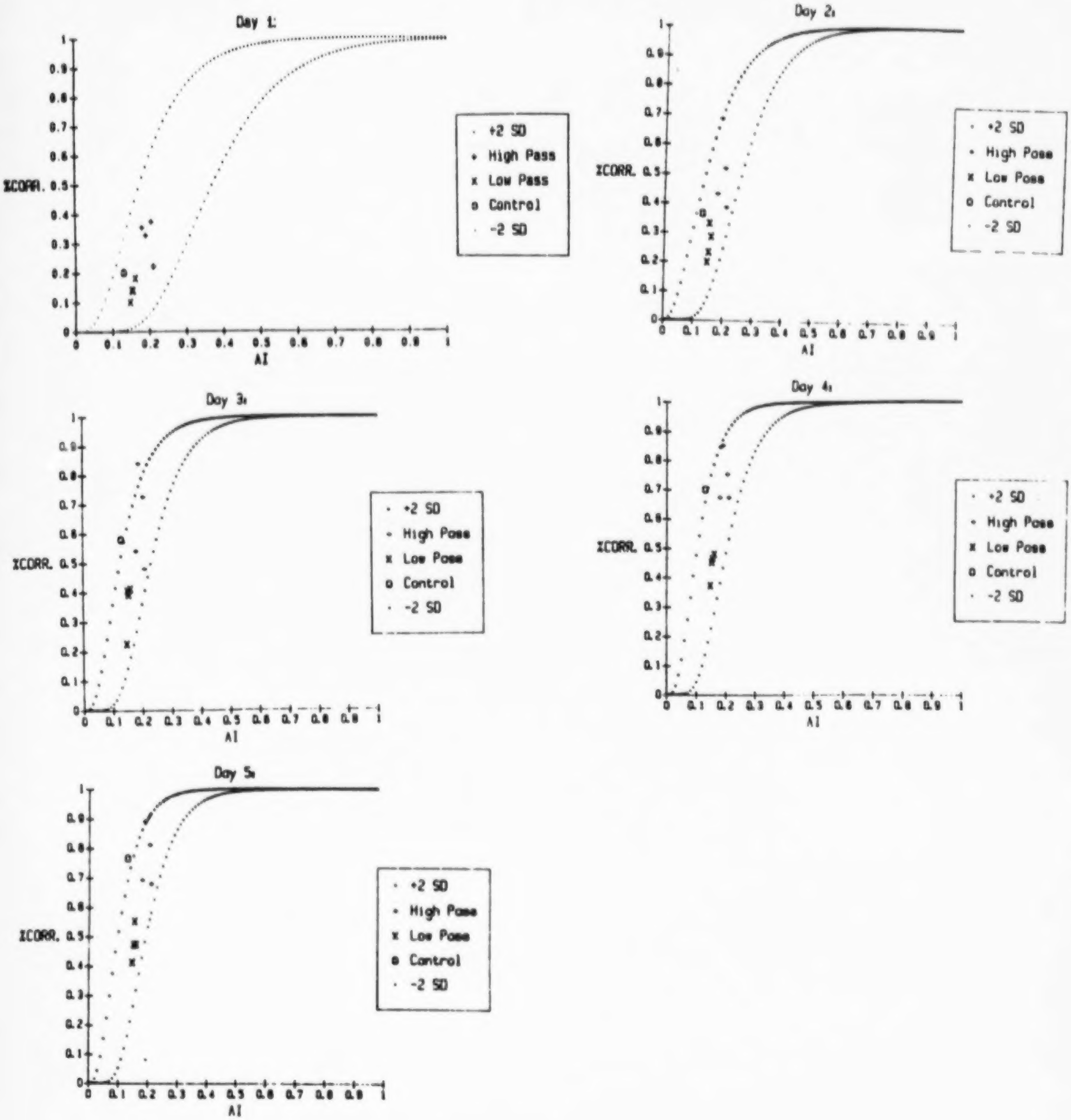


Figure 5a-e. Percent correct identification of filtered speech in Noise (0 dB S/N). Data points for high pass filtered speech are plotted as "+" lov pass filtered speech "x" and for the unfiltered control vith "o". The stop band attenuation levels (10, 20, 40 and 80 dB) are shown in parentheses next to each data point.

Discussion

These data provide some initial support for the hypothesis that the observed enhancement effect with spectral shaping is due to release from upward spread of masking in speech. The present results are also qualitatively similar to the earlier findings from studies by Nye et al. (1974) and Rand et al. (1973) which demonstrated a release from masking effect on the order of 20 dB in a dichotic listening task using highly controlled synthetic speech stimuli.

The initial motivation for this study was to examine the perceptual consequences of the change in the shape of the long term spectrum that takes place when the talker is speaking in noise. We know from earlier studies that when a speaker is talking in noise, there is a change in the long term spectrum similar to that produced by the manipulation used in the present study. We also know that speech produced in noise is more intelligible than speech produced in the quiet even when overall amplitude is controlled for (Dreher & O'Neill 1957; Summers et al. in press).

The present study has demonstrated that manipulating the shape of the long term spectrum enhances speech intelligibility in noise. The present results suggest that the underlying perceptual mechanism responsible for this enhancement is a release from masking effect from attenuation of the first formant. This finding is consistent with a hypothesis proposed by Lane and Tranel (1971) that speakers alter the characteristics of their speech, including overall amplitude, duration, mean fundamental frequency, and the shape of the long term spectrum, so as to maintain intelligibility in a noisy environment. Thus, the talker shows evidence of being sensitive to the demands on the listener in adverse listening conditions.

The present results also raise some questions about the validity of the assumption made in most formulations of the articulation index, that different frequency bands contribute independently to intelligibility. In one of the original studies on the articulation index, Fletcher and Galt (1950) discussed a model for speech-on-speech masking as one term in their formulation of the articulation index. This model is quite complex and it would not be appropriate to discuss it here in detail. In simple terms, however, their model is presented as a formula which enables the prediction of expected levels of masking based on the relative amounts of speech energy in different parts of the frequency spectrum. In general terms, the model predicts that frequency components of the speech signal with higher levels of energy will tend to mask lower energy components of the signal. The present study has shown significant interband effects on intelligibility which cannot be accounted for under the assumption of band independence. The observed performance in the high-pass filtered condition compared to the predictions assuming that there is no speech-on-speech masking provides some support for Fletcher and Galt's original proposal.

References

Dreher, J. & O'Neill, J. J. (1957). Effects of ambient noise on speaker intelligibility for words and phrases. *Journal of the Acoustical Society of America*, *29*, 1320-1323.

Flanagan, J. L. & Saslov, M. G. (1958). Pitch discrimination for synthetic vowels. *Journal of the Acoustical Society of America*, *30*, 435-442.

Fletcher, H. & Galt, R. H. (1950). The perception of speech and its relation to telephony. *Journal of the Acoustical Society of America*, *22*, 89-151.

French, N. R. & Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America*, *19*, 90-119.

Hirsh, I. J. (1948). The influence of interaural phase on interaural summation and inhibition. *Journal of the Acoustical Society of America*, *20*, 536-544.

Humes, L. E., Dirks, D., Bell, T. S., Ahlstrom, C., & Kincaid, G. E. (1986). Application of the articulation index and the speech transmission index to the recognition of speech by normal-hearing and hearing-impaired listeners. *Journal of speech and Hearing Research*, *29*, 447-462.

Kryter, K. D. (1946). Effects of ear protective devices on the intelligibility of speech in noise. *Journal of the Acoustical Society of America*, *18*, 413-417.

Lane, H. & Tranel B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, *14*, 677-709.

Licklider, J. C. R. & Miller G. A. (1951). The perception of speech. In S. Stevens (Ed.), *Handbook of experimental psychology*. New York: John Wiley and Sons.

Markel J, D. & Gray, M. (1976). *Linear prediction of speech*. New York: Springer-Verlag.

Nye, P. W., Nearey, T. M. & Rand T. C. (1974). Dichotic release from masking: further results from studies with synthetic speech stimuli. *Haskins laboratories status report on speech research SR-37/38*.

Rand, T. C. (1973). Dichotic release from masking for speech. *Haskins laboratories status report on speech research SR-33*, 47-55.

Summers, W. V., Pisoni, D. B. P., Bernacki, R., Pedlow, R. I., and Stokes M. A. (in press). Effects of noise on speech production: Acoustic and perceptual analyses. *Journal of the Acoustical Society of America*.

Thomas, I. B. & Niederjohn, R. J. (1968). Enhancement of speech intelligibility at high noise levels by filtering and clipping. Journal of the Audio Engineering Society, 16, 412-415.

Thomas, I. B. & Niederjohn, R. J. (1970). The intelligibility of filtered clipped speech in noise. Journal of the Audio Engineering Society, 18, 299-303.

Thomas, I. B. & Ohley, W. J. (1972). Intelligibility enhancement through spectral weighting. Proceedings of the 1972 IEEE conference on Speech Communication and Processing, 360-362.

Thomas, I. B. & Ravindram, A. (1971). Preprocessing of an already noisy speech signal. Journal of the Acoustical society of America, 133 (A).

Webster, J. C. & Klump, R. G. (1962). Effects of ambient noise and nearby talkers on a face to face communication task. Journal of the Acoustical Society of America, 34, 936-941.

Wegel, R. L. & Lane, C. E. (1924). The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. Physical Review, 23, 266-285.

[RESEARCH ON SPEECH PERCEPTION Progress Report No.13 (1987) Indiana University]

On the Arguments for Syllable-Internal Structure*

Stuart Davis

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington IN 47405

---

*This research was supported by an NIH Training Grant NS-07134-09 to Indiana University at Bloomington. I thank David Pisoni for his comments and discussion.

Abstract

Fudge (1987) examines evidence bearing on syllable-internal structure from speech errors, language games, distributional constraints, rhyming traditions, and languages that have an upper limit on the length of a syllable-final vowel-consonant sequence. Fudge argues against Clements & Keyser's (1983) contention that this evidence does not support the Rhyme as a syllable-internal constituent. He argues instead that this evidence unambiguously supports the division of the syllable into Onset and Rhyme. In this paper I argue that the speech error and word game data that Fudge cites in support of the Onset-Rhyme division do not provide evidence for the Rhyme as a syllale-internal constituent. They are compatible, though, with a syllable structure consisting of the constituents Onset, (syllable-initial consonant or consonants), Peak (vowel or other syllable peak), and Coda (syllable-final consonant or consonants). Moreover, I show that the other evidence that Fudge adduces in support of the Onset-Rhyme division actually are not relevant for determining constituency.

On the Arguments for Syllable-Internal Structure

Introduction

Fudge (1987) reconsiders five arguments that were made by Clements & Keyser (1983) to argue against the Rhyme (ie, the grouping of Peak and Coda) as a syllable-internal constituent. Fudge attempts to show that on closer scrutiny these five arguments actually support the Rhyme as a syllable-internal constituent. These arguments are based on evidence from speech errors, word games, distributional constraints, rhyming traditions, as well as from languages that have an upper limit on the length of a syllable-final vowel-consonant sequence. In this paper I will focus on the arguments from speech errors, word games, and distributional constraints and only briefly comment on the other two arguments. I show that these arguments do not in fact provide evidence for a syllable structure like that in Figure 1 that recognizes the Rhyme as a syllable-internal constituent, rather they are more compatible with a syllable structure like that in Figure 2 that only recognizes Onset, Peak, and Coda as subsyllabic units.

```
    Syllable                Syllable
       /\                     /|\
  Onset Rhyme            Onset Peak Coda
          /\
      Peak Coda

    Figure 1                Figure 2
```

Specifically, I will argue that the evidence from speech errors and word games that Fudge cites in support of the Onset-Rhyme division actually does not support the Rhyme as a syllable-internal constituent because the speech error and word game data Fudge considers only involves monosyllabic words, when polysyllabic words are considered it becomes obvious that a division is being made between the Onset and the rest of the word (not between Onset and Rhyme). Moreover, it will be argued that the pattern of exchange errors that occur in speech errors and the types of word games that involve movement of a part of a syllable or the insertion of a sequence within a syllable are compatible with a syllable having the structure in Figure 2. Finally, I will show that the argument from distributional constraints cited by Fudge for the structure in Figure 1 is really not relevant for determining syllable-internal constituency. Thus it is concluded that the arguments Fudge cites in support of the Rhyme as a syllable-internal constituent, as in Figure 1, are actually more compatible with the rhymeless structure in Figure 2 in which the Onset, the Peak and the Coda are the only subconstituents of the syllable.

Before considering Fudge's arguments from speech errors and word games in detail it is worth pointing out that, contrary to what Fudge (1987) states, Clements & Keyser (1983) do not argue for the syllable structure in Figure 2; they do not recognize Onsets and Codas as constituents of the syllable. This is clear from their following remarks:

> As far as we have been able to determine, there is no linguistic evidence suggesting that phonological rules ever make crucial reference to the categories "onset" and "coda". Thus, it appears that the set of syllable structure conditions defining the set of well-formed syllables for each language can be stated with complete adequacy with reference to the categories "syllable" and "nucleus". For example, the distinction

> between initial consonant clusters and final consonant clusters, which are subject to independent constraints, can be characterized directly with reference to the brackets which delimit the boundaries of the syllable [Clements & Keyser 1983:16].

Clements & Keyser recognize only the Nucleus (or Peak) as a subsyllabic constituent. The Onset and Coda do not have status as constituents in their view. The structure in Figure 2 is argued for most thoroughly in my dissertation [Davis 1985]. Let us now consider the speech error and word game phenomena that support the syllable structure in Figure 2.

The Argument from Speech Errors and Word Games

Speech errors

Based on the speech errors included in the Appendix of Fromkin (1973) Fudge contends that the fact that there are far more errors that exhibit Peak-Coda cohesiveness than Onset-Peak cohesiveness constitutes strong evidence for the reality of the Rhyme. Fudge (p. 372) cites the following three cases as errors that illustrate Peak-Coda cohesiveness:

(a) Spoonerisms such as "if the fap kits" for "if the cap fits" (Fromkin's example C. 20 (1973:245)):

(b) Haplologies such as "prodeption of speech" for "production and perception of speech" (O. 8, p.257);

(c) Blends such as "Irvine's quite clear" for "Irvine's quite near/close." (U. 27, p.261).

However, these speech errors actually do not support Peak-Coda cohesiveness; if anything the spoonerism example and blend example support the Onset as a constituent. Consider the spoonerism shown above. The fact that the syllable-initial consonant of 'cap' interchanges with the syllable-initial consonant of 'fit' can be taken as evidence for the constituency of the Onset, especially when it is pointed out that this type of speech error is very common. (See, for example, Shattuck-Hufnagel 1983). By no means, though, do such errors provide evidence for the (syllable-internal) constituency of the part of the word remaining after the interchange of syllable-initial consonants. This is obvious when we consider spoonerisms involving polysyllabic words like the following (taken from the Fromkin (1973) corpus):

(d) "heft lemisphere" for "left hemisphere" (C. 1, p. 245));

(e) "Yoman Rakobson" for "Roman Jakobson" (c. 7, p. 245).

In these examples there is cohesiveness of all the phonemes after the word-initial consonant. However, obviously, in 'hemisphere' and 'Jakobson' these phonemes cannot form a syllable-internal constituent since they span over more than a single syllable. Thus an example like that in (a) is just an instance of the cohesiveness of all phonemes after the initial consonant. What looks like Peak-Coda cohesiveness in the example that Fudge cites is really an instance of the cohesiveness of everything after the word-initial

Onset, and consequently, such examples are not relevant for determining the constituency of the Rhyme.

Furthermore, the cohesiveness of what does not move or interchange provides no evidence for constituency, in general. This point becomes apparent by making the analogy with syntactic movement phenomena that has been used to argue for syntactic constituency. Consider the following sentence in (f) and the corresponding cleft construction in (g)

(f) The man put the book on the table.

(g) It was on the table that the man put the book.

Just because the phrase "on the table" moves in forming the cleft construction does not mean that what remains "the man put the book" forms a constituent. The evidence for constituent structure comes from what moves not from what remains behind after movement has taken place. Thus spoonerisms like those in (a), (d), and (e) cannot provide evidence for (or against) the Rhyme. If these spoonerisms provide evidence for a syllable-internal constituent, it is the Onset that they provide evidence for since it is the syllable-initial consonants that are exchanged in each of the errors.

The identical argument against the alleged Peak-Coda cohesiveness in these errors is relevant with the blend in (c) where what looks to be partaking in the blend is the syllable-initial consonants of 'close' with the vowel and syllable-final consonant of 'near'. This is taken by Fudge to be an example of Peak-Coda integrity and thus evidence for the Rhyme. However, if we focus on blends involving polysyllabic words it becomes quite evident that syllable-internal constituency has no role in the formation of blends. Consider the following three blends taken from Fromkin (1973):

(h) "scalary" for "scale/salary" (U. 12, p.260

(i) "adjoicent" for "adjoining/adjacent" (U. 39, p.261);

(j) "recoflect" for "recognize/reflect" (U. 62, p.261).

As the example in (h) shows blends can combine syllable-initial consonants of one word with the remainder of the second word. The blend Fudge cites in (c) is an instance of this. The examples in (i) and (j) show that blends do not necessarily combine the Onset of one word with the remainder of the other, rather the division point between the two words can vary. It could be after a syllable boundary as in (i) or after a Peak as in (j). These examples show that the first part of the blend and the second part of the blend do not necessarily form syllable-internal constituents. That examples like (h), (i), and (j) are fairly common suggests that blends do not in fact support the constituency of the Rhyme.

Finally, it is unclear how the haplology error that Fudge cites from Fromkin (1973) shown in (b) argues for Peak-Coda cohesiveness. If "prodeption" is a haplology from 'production and perception' obviously the part that has deleted ("uction and perc") cannot be a syllable-internal constituent, and the parts that have remained ('prod' and 'eption') do not comprise syllable-internal constituents either. Thus this piece of evidence does not support the Rhyme.

It has been shown that the type of speech error evidence Fudge used to support the constituency of the Rhyme in fact does not really bear on it. There is speech error evidence, though, that can provide evidence for syllable-internal structure. These are the spoonerisms like in (a). However, as discussed earlier, what is crucial for determining constituency is what moves in these errors and not what remains behind. In this way, constituency is determined on analogy with movement phenomena in syntax. When analyzing the patterns of spoonerisms that occur in speech errors what one finds (and this has been pointed out by Shattuck-Hufnagel 1983) is that in the great majority of cases syllable-initial consonants interchange with syllable-initial consonants, vowels with vowels, and syllable-final consonants with syllable-final consonants. That these types of spoonerisms are most frequent is quite compatible with the view that the syllable consists of Onset, Peak, and Coda. Spoonerisms that involve a consonant-vowel sequence of one word interchanging with a consonant-vowel sequence of another word or involve a vowel-consonant sequence of one word interchanging with a vowel-consonant sequence of another word are rare. (Only 8% of all spoonerisms in the large speech error corpus reported on by Shattuck-Hufnagel 1983.) In discussing what kind of syllable structure these speech errors are most compatible with Shattuck-Hufnagel (1983:117) states the following:

> [T]he hypothesis that the syllable onset, nucleus, and coda are the primary units of sublexical serial misordering accounts for a higher proportion of sublexical exchange errors than does the single-segment hypothesis [ie, that there is no syllable-internal constituent structure] or the onset and rhyme hypothesis.

We take her findings then as support for the structure in Figure 2. The speech error data cited by Fudge does not provide evidence for the constituency of the Rhyme.

Word Games

Fudge (1987:373) argues that word games in which parts of syllables are moved, deleted, or broken up provide evidence for the Onset-Rhyme division. Once again Fudge interprets as an Onset-Rhyme division what is really a division between Onset and the rest of the word. Fudge cites Cockney rhyming slang as providing evidence for Peak-Coda cohesiveness. He cites the example "apples and pears" for 'stairs' (in which the /p/ of 'pears' replaces the deleted /st/ of 'stairs'). However, when other examples are considered it becomes clear that the cohesiveness involves everything after the word-initial Onset. This is made evident by the bisyllabic examples like "Derby Kelly" for 'belly' where what remains identical in the affected word is everything after the Onset. What looks like Peak-Coda cohesiveness in the example Fudge cites is only really a case of cohesiveness of everything after the word-initial Onset and thus has no bearing on the question of syllable-internal structure. Rhyming slang actually provides evidence for the constituency of the Onset since one syllable-initial cluster replaces another.

Moreover, contrary to what Fudge states (1987:373), the Pig Latin word game does not imply a major split between Onset and Rhyme. It may seem so from the example he gives (street ---> eetstrey) in which there is Peak-Coda cohesiveness while the Onset moves. However, it is quite obvious from considering the Pig Latin forms of polysyllabic words that the major split is between Onset and the remainder of the word, not between Onset and Rhyme. This is made clear by such examples as 'Latin' ---> "atinley" and 'criminal'

---> "iminalcrey" in which there is cohesiveness of everything after the word-initial Onset. What looks to be an Onset-Rhyme split based on monosyllabic examples (like 'street' ---> "eetstrey") is really just an instance of a split between Onset and the rest of the word. Pig Latin has no bearing on the constituency of the Rhyme. It does, though, provide evidence for the constituency of the Onset since syllable-initial consonants move as a unit.

Fudge also cites the "op" word game where the sequence /ap/ is inserted before the vowel of every syllable (eg, 'give' ---> /gapIv/, 'robin' ---> /rapabapIn/ in which the inserted sequences are underlined) as evidence for the Onset-Rhyme division. I show instead that the "op" word game provides evidence for the Onset-Peak division and that word games which insert a sequence of phonemes within the syllable support the syllable structure in Figure 2. First, note that the "op" word game can be interpreted as involving insertion between either the Onset and Rhyme or the Onset and Peak. Now, consider the English word game cited by Laycock (1972:74). In this word game the sequence -gV (in which V stands for a copy of the preceding vowel) is inserted after every vowel. Hence, in this word game 'pin' is pronounced as /pIgIn/. This game seems to split up the Rhyme since insertion occurs between the Peak and the Coda. A similar type of word game is reported for Spanish in Sherzer (1982) in which the sequence -fV is inserted after every vowel. For example, the word 'grande' has the word game form /grafandefe/. Finally, Chinese has a word game described in Yip (1982) that inserts the sequence -ayk after the Onset. For example, the word 'pey' has the game form paykey. This game can be interpreted as having insertion between Onset and Rhyme or between Onset and Peak. What these word games (and other word games like them--see, for example the survey of word games in Laycock (1972)) that involve insertion within a syllable show is that if the inserted sequence begins with a vowel it is inserted between the Onset and the Peak (as in the "op" game and the Chinese word game), and if the inserted sequence begins with a consonant it is inserted between the Peak and the Coda (as in the English -gV insertion word game cited by Laycock and the Spanish -fV insertion word game cited by Sherzer). Word games like those just cited (as well as others cited in Laycock 1972) that involve insertion within the syllable do not seem to break up Onsets, Peaks, or Codas. That such games respect the integrity of these is taken as providing evidence for the syllable structure in Figure 2. These types of games do not provide evidence for the Onset-Rhyme division.

Finally, as Fudge points out, there are word games that involve the interchange of syllable-final VC-sequences, as in Burmese. This type of game can be taken as providing evidence for the constituency of the Rhyme. Fudge also points out that there are word games like that in Hanunoo which involve an interchange of syllable-initial CV-sequences. This type of game can be taken as evidence against the constituency of the Rhyme. However, judging from the typology of language games provided in Laycock (1972) it seems that word games like that mentioned for Burmese and Hanunoo are extremely rare. If word games provided evidence for the constituency of the Rhyme it would be expected that word games like that in Burmese would be common and ones like in Hanunoo nonoccurring. If syllable structure is as in Figure 2, though, it would be expected that these two types of word games would be extremely rare since two constituents are involved (the Peak and the Coda in the Burmese game and the Onset and the Peak in the Hanunoo game). That word games like these seem to be extremely rare can be taken as supporting the syllable structure in Figure 2. Thus, the evidence from word games that Fudge cites in support of the constituency of the Rhyme do not actually support it. They do, though, provide evidence for the syllable structure in Figure 2.

The Argument from Distributional Constraints

Fudge (1987) argues against Clements & Keyser's (1983) position that the common occurrence of distributional constraints between parts of the syllable other than between Peak and Coda provide evidence against the Rhyme. Fudge shows that the majority of Onset-Peak constraints that Clements & Keyser cite for English can very easily be accidental. As for constraints between Onset and Coda, Fudge (1987:369) dismisses these as being irrelevant for the status of the Rhyme. Finally, Fudge cites a number of general constraints that hold between Peak and Coda and argues that these provide evidence for the Rhyme. In this section, however, I argue that the constraints between Onset and Coda in English cannot be dismissed so readily as having no bearing on the Rhyme (assuming that the existence of distributional constraints between two items is a legitimate test for their comprising a constituent). But I also contend that, in general, cooccurrence constraints are not a good diagnostic for constituency.

English has a number of cooccurrence constraints that hold between Onset and Coda. For example, in monosyllables of the form sCVC the same noncoronal consonant cannot flank both sides of the vowel nor can nasal consonants flank both sides of the vowel. These constraints are exceptionless. Fudge states (1987:369) that Onset-Coda constraints "...do not reflect constraints holding between Onset and Peak or Onset and Rhyme, but constraints between Onset and Coda, and are therefore irrelevant to the status of Rhyme." Technically, this is correct. It would seem, though, that if constraints between Peak and Coda provide evidence for the constituency of the Rhyme as Fudge argues, constraints holding between Onset and Coda should provide evidence that Onset and Coda together form a constituent. Hence, the Coda would form one constituent with the Peak (based on the existence of Peak-Coda cooccurrence constraints) and another constituent with the Onset (based on the existence of Onset-Coda cooccurrence constraints). Moreover, if at least some of Clements & Keyser's (1983) proposed constraints between Onset and Peak in English are indeed correct then the Peak would also form a constituent with Onset. Thus, if syllable-internal structure is based on the existence of cooccurrence restrictions, syllable structure would be as in Figure 3 in which there is "double motherhood" for each of Onset, Peak, and Coda.

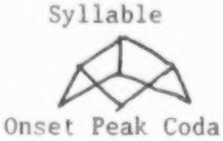


Figure 3

The implausibility of the structure in Figure 3 leads to the conclusion that the existence of cooccurrence constraints between two elements does not provide evidence that those two elements comprise a single constituent. To this end, it is worth pointing out that in syntax cooccurrence restrictions do not necessarily indicate constituency. For example, subject verb agreement in English fails to establish that subject-NP and verb form a single constituent, while, in Spanish, subject-NP and predicate-adjective agreement for number and gender fail to establish that subject-NP and predicate adjective comprise a single constituent. Hence, the existence of cooccurrence restrictions cannot be used to establish the constituency of Rhyme or any other syllable-internal constituent. (It should also be pointed out that cooccurrence constraints between elements in Onset and Coda as well as between elements in Onset and

Peak exist in many languages. In addition, there are languages which display distributional constraints between phonemes not within the same syllable. Such languages, which are discussed more fully in Davis (1985), provide further evidence that the existence of cooccurrence constraints is not a diagnostic for syllable-internal constituency.)

Other Arguments

Fudge (1987) briefly discusses two other possible arguments for the constituency of the Rhyme that were mentioned by Clements & Keyser (1983). One argument relates the poetic notion 'rhyme' with the linguistic notion 'Rhyme'. Both Fudge and Clements & Keyser correctly observe that these notions are not identical. The linguistic notion 'Rhyme' refers to the part of the syllable after Onset while the poetic notion 'rhyme' refers to everything after the stressed vowel in a word. Thus poetic rhymes often do not coincide with the linguistic notion 'Rhyme', as examples like "sinister-minister" or "seventeen-Levantine" attest. Consequently, the poetic notion 'rhyme' really does not bear on the status of the linguistic 'Rhyme' as a syllable-internal constituent.

The second argument relates to languages that impose an upper limit on the length of the Rhyme. Fudge (1987:371) notes that the crucial factor of this argument is really not the length of the rhyme but the weight of the Rhyme. He further states (p. 371):

> [I]n the great majority of languages in which concepts of syllable weight are defined, they are defined in terms of Rhyme rather than of any other element or combination of elements within the syllable. In particular, the Onset appears to have no part to play in constraining syllable weight[.]

While it is no doubt true that there are many languages in which either the presence of a long vowel or a consonant in the Coda makes a syllable heavy, it seems that in the majority of languages syllable weight is either not relevant (judging from the survey of stress systems in Hyman 1977) or only a property of the Peak (ie, only a long vowel makes a syllable heavy and not the presence of a consonant in the Coda). Thus, for these languages the argument for syllable weight cannot be made for the constituency of the Rhyme. Furthermore, there are some cases in which the Onset does appear to play a part in constraining syllable weight. These are discussed in more detail in Davis (to appear). Moreover, just because a rule (such as a stress rule) might make reference to syllable weight does not necessarily argue that Peak and Coda comprise a single constituent. Other types of rules have other environments, and the elements mentioned in rule environments do not have to comprise constituents. Hence, the argument from syllable weight does not provide conclusive evidence for the syllable structure in Figure 1.

Conclusion

In this paper I have shown that the evidence which Fudge cites to support the syllable structure of Figure 2 does not actually support it. The evidence from speech errors and word games are more compatible with the rhymeless syllable structure of Figure 2 than it is with the syllable structure of Figure 1, while the evidence from distributional constraints, rhyming traditions, and syllable weight do not bear on the question of syllable-internal constituency. Thus, I would disagree with the conclusion of Fudge (1987:376) that such evidence leads us to support the structure in Figure 1 as the best model of the syllable. It could be that Figure 1 is the best model of the syllable, but the evidence that Fudge cites does not show it. What the evidence from speech errors and word games do show, however, is that the syllable consists of at least Onset, Peak, and Coda.

References

Clements, N. & Keyser, J. (1983). *CV phonology*. Cambridge: MIT Press.

Davis, S. (1985). *Topics in syllable geometry*. Doctoral dissertation, University of Arizona, Tucson.

Davis, S. (To appear). Syllable onsets as a factor in stress rules. *Phonology*, *5.1*.

Fromkin, V. (ed.) (1973). *Speech errors as linguistic evidence*. The Hague: Mouton.

Fudge, E. (1987). Branching structure within the syllable. *Journal of Linguistics*, *23*, 359-377.

Hyman, L. (1977). On the nature of linguistic stress. In L. Hyman (ed.), *Studies in stress and accent* (pp. 37-82). Los Angeles: Department of Linguistics, University of Southern California.

Laycock, D. (1972). Towards a typology of ludlings or play-languages. *Linguistic Communications*, *6*, 61-113.

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. MacNeilage (ed.), *The production of speech* (pp. 109-136). New York: Springer-Verlag.

Sherzer, J. (1982). Play-languages: with a note on ritual languages. In L. Obler and L. Menn (eds.), *Exceptional language and linguistics* (pp. 175-199). New York: Academic Press.

Yip, M. (1982). Reduplication and C-V skeleta in Chinese secret languages. *Linguistic Inquiry*, *13*. 637-661.

[RESEARCH ON SPEECH PERCEPTION Progress Report No. 13 (1987) Indiana University]

The Identification of Speech Using Word and Phoneme Labels*

John Logan
Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, IN 47405

---

*I thank David Pisoni for his advice and commments, Van Summers for providing his "bad-pad" stimuli, and Mike Stokes for his help in testing subjects. This research was supported, in part, by NIH Research Grant NS-12179-11 to Indiana University in Bloomington.

Abstract

The purpose of the present investigation was to see under what conditions pre-lexical and phonological knowledge are used by listeners to identify speech. Listeners identified a continuum of synthetic CVC stimuli using either a label that corresponded to the entire syllable or to the initial consonant alone. In Experiment 1, subjects heard stimuli corresponding to words. Results supported the prediction that response times (RT) would be faster for word labels than for phoneme labels due the facilitory effect of phonological information from the listener's lexicon. Experiment 2 was similar to Experiment 1 but also included a condition in which listeners identified a continuum of nonword CVC stimuli using labels corresponding to the orthographic representation of the syllable as well as phoneme labels corresponding to the initial phoneme of the nonword syllable. For the word stimuli, the findings of Experiment 1 were replicated. In contrast, listeners identified the nonword stimuli faster when they used phoneme labels than when they used word-like labels. The results indicated that word and phoneme labels elicit different types of responses depending upon the nature of the stimulus. The RT advantage of word labels for identifying spoken words is explained in terms of lexical activation and access to phonological information whereas the RT advantage of phoneme labels for identifying spoken nonwords is explained as the consequence of attention being directed to a pre-lexical level of perceptual analysis.

The Identification of Speech Using Word and Phoneme Labels

The internal representation of speech has been the subject of much speculation. Linguistics has provided a representational hierarchy for spoken language beginning with features and continuing with phonemes, syllables, and words (morphemes). There are higher order representations but word units are generally thought to be the smallest representational unit by which lexical access occurs. Lexical access is important because it is the step in the perceptual process where semantic information makes contact with the content-less representations of earlier stages of processing. By content-less representations, I refer to the lack of semantic content found in hypothetical representations such as features, phonemes, and syllables. The steps leading to lexical access thus form a natural area of study delimited by the conjunction of perceptual events and semantic knowledge. In the present paper, two experiments to investigate the nature of the representational units used for lexical access and the effect of lexical information on the identification of speech are described.

Psychological studies of the perceptual representations of speech are primarily based on linguistic descriptions of language. As an example, a major goal for psychologists studying speech perception has been to understand how acoustic cues in the speech signal are used by listeners to generate phonemic representations (Pisoni & Luce, 1987). Much of this work used simple CV (consonant-vowel) syllables such as /ba/ and /da/ as stimuli. The use of CV stimuli was likely due to their relatively simple structure and also the kinds of questions that researchers were asking. Researchers were interested in phoneme perception because phonemes were assumed to be the simplest units that linguistic description posited. And, it followed logically that if the mechanisms responsible for phoneme perception could be described, that this information would provide the basis for understanding the perception of larger linguistic units such as words. In short, linguistic units and their hierarchical relationship were equated with perception by psychologists, a principle that has guided much of the research that psycholinguists have done over the last thirty-five years.

The synthesis of word or syllable units from component phonemes is a bottom-up process. However, another possibility exists. That is, the internal units comprising words, phonemes, may not be used in the perceptual process itself but instead may be a consequence of some post-perceptual process (Savin & Bever, 1970). Within this framework, phonemes are considered to be a by-product of the analysis of words or syllable-level units into their component parts rather than being the precursors of the word units. Acoustic information from the speech signal would therefore make contact directly with higher level lexical representations without an intermediate phonemic representation. This view is embodied in Klatt's (1979) LAFS model in which lexical information is accessed directly from spectral information in the signal and phonemes are derived from the knowledge associated with the lexical unit.

A third view of the relationship between prelexical and lexical processes is that both kinds of processes are operational during speech perception. Within this framework, several variations have been proposed. One variation posits that prelexical and postlexical processes are in competition and operate more or less simultaneously. Under some conditions, one or the other process is favored. The process that is favored finishes first and is the one that determines a subject's response time (RT) for whatever task the subject is doing. The dual-code hypothesis of Foss and Blank (1980) is an example of such a model as is Cutler and Norris's (1979) "race model." Dell and Newman

(1980) also proposed a model closely patterned after Foss and Blank's model. A second way to view the relationship between prelexical and postlexical processes is as an interactive process in which sensory information interacts with lexical information, each affecting the other. Proponents of this view include McClelland and Elman (1986) and Marslen-Wilson and Welsh (1978). Each of the theoretical positions outlined above has some empirical support which I will describe below.

Evidence for the derivation of prelexical units from larger units comes from several sources. Savin and Bever (1970) used a monitoring task in which listeners were presented lists of nonsense syllables. Some subjects were given an initial phoneme to detect while other listeners were given an entire syllable to detect. The former was called a phoneme-monitoring condition while the latter was called a syllable-monitoring condition. Subject's RTs were significantly shorter in the syllable-monitoring condition than in the phoneme-monitoring condition. Savin and Bever interpreted their results as indicating that phonemes were preceived only as a consequence of an analysis of syllable-sized units. That is, although phonemes were psychologically real, syllables were the initial unit of perception and phonemes were derived from a decomposition of information contained in syllables.

Savin and Bever's interpretation of their results was criticized by McNeill and Lindig (1973). They considered the results of Savin and Bever to be the result of a mismatch between the representational level of the target and the level of the items for which subjects were required to monitor. In other words, listening for phonemes in the context of syllables required more response time than listening for syllables in the context of syllables because of the incompatability of the two levels of representation in the phoneme-syllable case. McNeill and Lindig carried out a study in which they examined all possible combinations of target and list items among phonemes, syllables, words, and sentences in a monitoring task. Their results showed that minimal response times could be found at any linguistic level if there was strict compatability between target and search list. McNeill and Lindig suggested that incompatability between stimulus and target caused detrimental effects in monitoring experiments because listeners have difficulty in allocating attention between different perceptual levels simultaneously. Furthemore, they argued that the results of monitoring experiments of the type Savin and Bever conducted (eg., Foss & Swinney, 1973), were incapable of providing answers to questions regarding the primacy of one perceptual level versus another. In fact, McNeill and Lindig argued that there is "...a series (or network) of processing stages and each can in principle be the focus of attention." (p. 430). The results of McNeill and Lindig were later confirmed by Healy and Cutting (1976).

Rubin, Turvey, and van Gelder (1976) introduced an additional variable to the phoneme monitoring task, the lexical status of the target item. Noting that subjects can detect letters faster in printed words than in nonwords, Rubin et al. carried out a similar experiment using spoken words and nonwords. They found that initial phonemes were detected faster in words than in nonwords. Thus, the lexicality of a item affected the ability of subjects to detect target phonemes embedded within the item. This finding seemed to indicate that lexical information could facilitate decisions about the presence or absence of a target phoneme in a word. On the other hand, no such facilitory effect would be observed for nonwords due to the lack of lexical information for such items. They interpreted this result as showing that a word was more available to consciousness due to the greater familiarity of listeners with lexical information than with nonlexical information. Thus, in Rubin et al.'s experiment, the mismatch between phoneme and target word was not the critical variable affecting the results since both word and nonword stimuli were the same-sized units. Instead, the meaningfulness and lexical status of the stimulus item within which the target phoneme was embedded was the critical factor.

Additional evidence has also been collected by other investigators that also suggested the lexicality of an item determined subjects' ability to detect phonemes within that item. Morton and Long (1976) varied the transitional probability of an item's occurrence in a phoneme monitoring task and found that the predictability of the word containing the target phoneme was positively correlated with the RT for the detection of the target phoneme. They argued that their results indicated that lexical access occurs prior to phoneme identification. However, later work suggested that both top-down and bottom-up effects could mediate the detection of phonemes in words.

In two experiments similar to Morton and Long's experiment, Dell and Newman (1980) varied low-level phonetic features in addition to the predictability of the context preceding the target word, thus manipulating both top-down and bottom-up information. They found that phoneme detection was affected by the immediately preceding phonetic context as well the predictability of the sentence. If a word containing a phoneme differing by only one feature from the target phoneme was presented before the word containing the target phoneme in the sentence, the RT for the detection response was increased compared to the RT for a phonetically less-similar but synonomous word. Dell and Newman used their results to formulate a two-component model of how listeners make decisions about what phonemes are present within words. They proposed that if context is sufficient, top-down lexical information can be used to determine the presence of a target phoneme within a word. On the other hand, if the preceding context does not supply sufficient information, the detection decision is based on an analysis of prelexical information. RTs for decisions based on lexical information will tend to be faster when context activates the appropriate lexical entry. However, RTs for decisions based on prelexical information will be faster when insufficient context is present.

Foss and Blank (1980) proposed a similar model which they called the "dual-code" model. In Foss and Blank's model, both a prelexical phonetic code and a postlexical phonological code are present; factors related to the stimulus and task determine which code will be responsible for the subject's response. Under most conditions, the postlexical code is the one that determines access to the internal structure of words. However, when the signal-to-noise ratio is high and sufficient processing resources can be devoted to the task (a phoneme monitoring task, for example), subjects will be able to use the prelexical code to initiate a response.

In a follow-up to this work, Foss and Gernsbacher (1983) found that several of the predictions of the dual-code model were not substantiated when subjected to empirical testing. First, they found that low-level phonetic manipulations affected subjects RTs even when they were carrying out a demanding task that presumably encouraged the use of a postlexical code. Second, they argued that the lexical effects observed in earlier work, such as Foss and Blank (1980) and Rubin et al. (1976), were due to the uncontrolled effects of phonetic factors, primarily vowel length, in the target words. In short, Foss and Gernsbacher basically rejected the dual-code model and instead decided the evidence supported a prelexical model in which only bottom-up information was used to make judgements about the phonemic composition of a word.

Not all researchers were convinced that Foss and Gernsbacher's evidence condeming two-process pre- and postlexcial models was convincing. In a recent study, Cutler, Mehler, Norris, and Segui (1987) suggested that some of the low-level phonetic differences claimed by Foss and Gernsbacher to account for word-nonword differences in monitoring tasks were artifactual, and were due to differences in dialect between the subjects and materials used by Foss and Gernsbacher (Southern U.S.) and the source of the information used to back their claim about vowel length (Northern U.S., Peterson & Lehiste, 1960). Furthermore, Cutler et al. noted that despite Foss and Gernsbacher's claims, some studies, such as Rubin et al. (1976), found differences between words and nonwords under conditions in which differences between vowels were controlled.

Cutler et al. (1987) considered the claims regarding the relationship between prelexical and postlexical processes in the phoneme monitoring task to be very closely tied to the nature of the task and the stimulus materials used. They carried out a series of studies designed to establish under what circumstances pre- and postlexical processes would account for subject's phoneme monitoring responses. Although other factors such as the phonetic structure of the stimulus materials and the predictability of the context were implicated as contributing to the pattern of results observed, the effect of task monotony was considered by Cutler et al. as one of the primary reasons for the many inconsistent findings within the phoneme monitoring literature. They noted that, in general, experiments showing no lexical effect (that is, no difference between RT for detecting phoneme targets in words versus nonwords) contained monosyllabic targets whereas for experiments in which a lexical effect was obtained, the targets were included in lists with words containing more complex syllabic structures or even sentences. Empirical support for this conclusion was obtained when Cutler et al. repeated an experimental procedure which had previously resulted in a lexical effect. All bisyllabic words from the lists were removed and replaced with CVC words. Thus, the only difference between the two experimental procedures was the absence of the bisyllabic words in the second experiment. The results indicated that decreasing the variability of the stimuli eliminated the lexical effect that had been obtained in the earlier experiment. Cutler et al. viewed attention as the underlying mechanism responsible for the effect of stimulus monotony on the phoneme monitoring task. They predicted that whenever a homogenous list of stimuli is presented to subjects, the likelihood that no lexical effect would be observed is increased.

The phoneme monitoring task could be viewed as a task in which attention must be shared between lexical and phonetic levels of processing. If the list items become monotonous to the listener, the listener stops hearing the stimuli as meaningful speech and attention is likely shifted to the phonetic level, attenuating lexical effects. Cutler et al. suggested that if the lexical items are more varied, then attention remains at the lexical level. They cited the work of Samuel and Ressler (1986) as supporting their attention-based explanation. Samuel and Ressler used a phoneme restoration paradigm in which the subject's task was to discriminate between trials in which noise replaced phoneme and trials in which noise was added to mask a phoneme within a word. Subjects could not reliably discriminate between the two types of trials unless they were informed which phoneme in the word was going to be subjected to the noise manipulation. Thus, subjects could attend to low-level acoustic differences that were generally ignored if attention was directed to the appropriate region of the word. Their finding demonstrated that subjects tended to focus their attention on the lexical level when listening to speech but if required, they could focus on prelexical levels as well. The results of Samuel and Ressler provide some empirical support for

McNeill and Lindig's (1973) speculation that selective attention could, in principle, be shifted to various levels of linguistic processing.

Cutler et al. (1987) used the results of their experiments to dismiss the claims of Foss and Gernsbacher (1983) that bottom-up processes were the primary means by which the identification of phonemes within words occurred. Instead, Cutler et al. demonstrated that lexical effects could be obtained even when the low-level acoustic factors Foss and Gernsbacher claimed were responsible for the lexical effects were controlled for. They also reviewed several other models of spoken word recognition and considered how easily each could account for the pattern of their results, especially how attention could be implemented. Cutler et al. noted that the race model of Cutler and Norris (1979) had response outlets at separate levels which would allow responses to be the function of attention shifting between these different levels. In contrast, the TRACE model of McClelland and Elman (1986) would require an attentional mechanism to be added on to the already-existing interactive framework, a less "elegant" mechanism than found in the race model, according to Cutler et al. The dual code model of Foss and Blank (1980) was also compared with the race model of Cutler and Norris (1979). Although very similar, Cutler et al. argued that the race model predicted certain findings that the dual code model did not. One result predicted by the race model was that the occurence of lexical effects would be correlated with the length of the vowel following the phoneme target. This result was predicted on the grounds that increasing the length of the vowel, and thus the length of the word for CVC stimuli, "increases lexical access time and hence decreases the likelihood of the lexical output response winning the race" (Cutler et al., 1987, p. 170). Therefore, stimulus items with short durations would have a greater probability of being responded to via a lexical outlet and not the prelexical outlet. Cutler et al. analyzed their data and found that, indeed, lexical effects were more likely to occur in shorter stimulus items.

Other tasks have also been used to explore the nature of the relationship between pre- and postlexical processes and how information about the phonological structure of a word is obtained. In an influential study, Ganong (1980) examined how lexical information affects the phonetic categorization of speech. In Ganong's original experiment, he was interested in determining the locus of the effect of lexical knowledge in speech perception. That is, would lexical information affect the interpretation of low-level acoustic-phonetic information or would lexical information alter the interpretation of already-categorized prelexical units? Ganong constructed several synthetic speech continua in which one endpoint was a word and the other endpoint was a nonword. He found that subjects tended to classify the ambiguous stimuli from the middle of the continua as belonging to the word category rather than the nonword category. In order to choose between the two accounts of how lexical information could influence phonetic categorization, he examined the shape of the identification function for each continua. Based on the shapes of these functions, Ganong concluded that lexical information influenced the interpretation of low-level phonetic information, thus supporting an interactive model of spoken word perception.

Fox (1984) extended Ganong's work by measuring RTs in an identification task similar to the one used by Ganong. He found that the effect of lexical status was less pronounced at shorter RTs than at longer RTs. That is, when subjects latencies were divided into three ranges (slow, medium, and fast), the fast RTs were associated with a phoneme boundary near the center of the continuum whereas medium and slow RTs corresponded to an identification function that was shifted towards the nonword-end of the continuum, just as Ganong had obtained. Fox's results indicated that the lexical status of a word does not effect phonetic categorization until some measurable period of time has elapsed. That is, initially, no lexical effect on phonetic categorization can be observed; only after some period of time does lexical information appear to affect the identification of speech.

Connine and Clifton (1987) pursued the issue of how lexical information affects speech perception using the methodology developed by Ganong (1980) and Fox (1984). In their first experiment, word-nonword continua were used as stimuli. Subjects were asked to identify the stimuli and RT was measured. Connine and Clifton found that RTs for ambiguous stimuli were faster if they were identified as words. However, the mean RTs for unambiguous endpoint stimuli that were identified as words were no faster than for unambiguous nonword stimuli. In a second experiment, one important variation was added to the procedure: subjects were given different monetary payoffs for different category labels in an identification task using word-nonword stimuli. The purpose of the payoff manipulation was to introduce a post-perceptual bias to see if a different pattern of RTs could be obtained at the category boundaries compared to the endpoint stimuli. The results of Connine and Clifton's second experiment showed that subjects were biased to respond using one category label more than the other category label. However, in contrast to the results of the first experiment, the RTs for stimuli from the boundary region of the continua did not differ whereas RTs for endpoint stimuli that were consistent with the bias differed. Thus, the results of Experiment 1 suggested the operation of a perceptual mechanism whereas the results of Experiment 2 suggested the operation of a post-perceptual mechanism. Connine and Clifton viewed their results as evidence for an interactive speech perception mechanism that is separate from the mechanism responsible for effects such as the payoff bias observed in Experiment 2.

The relationship between pre- and postlexical information and how each is derived is an obvious candidate for research from a developmental perspective. An example of such an approach is a study carried out by Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, and Mehler (1988) who tested the ability of newborns and 2-month-old infants to detect phonetic differences between syllables. They used a procedure that required the infants to use perceptual representations of speech rather than simply tapping their ability to perform a same-different discrimination task that could be accomplished using sensory information alone. Bertoncini et al. found that the infants did not show evidence that the representation of speech sounds was based on phoneme categories. Instead, Bertoncini et al. argued that their results supported the existence of a holistic syllable-based level of representation in infants in which phonemes remain undifferentiated. Furthermore, they noted a developmental trend even within the limited age range of their subjects. The newborns tended to use a more global form of representation than the two-month-old infants. Other developmental studies have also shown that children have limited knowledge of the internal phonetic structure of words and syllables (eg., Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977), suggesting that although children may be good at discriminating speech stimuli based on acoustic-phonetic information, their internal represention of speech are not so well-differentiated. Some researchers suggest that even in adults, the analysis of speech at a phonemic or phonetic level only occurs under limited conditions, such as when learning a new language or listening to speech under degraded conditions (Cutler et al., 1987).

Taken together the results of phoneme monitoring experiments, identification experiments, and developmental studies suggest that either pre-lexical or post-lexical representations are available to subjects if information about the internal structure of speech is required. The source of this information depends on the nature of the task, the stimulus materials, and the level at which attention is focused. Although attention is generally focused at a lexical level, access to prelexical representations is possible. The present study was designed to examine some of these variables and assess how they affect the identification of speech. Two experiments were carried out, each using a task in which listeners identified synthesized speech stimuli using two kinds of labels. One set of labels consisted of the orthgraphic representation of the entire speech stimulus while the other set consisted of the orthographic representation of only the initial consonant. In both experiments, RT and identification data were collected.

Experiment 1

The purpose of the first experiment was to see the extent to which response labels could influence categorization judgements of speech stimuli modelled after monosyllabic English words. One group of subjects used labels corresponding to the orthographic representation of the entire word (eg., /ret/ = "rate") while another group used labels corresponding to the orthographic representaion of the initial phoneme (eg., /ret/ = "r"). Two continua were used, a RATE-LATE continuum and a BAD-PAD continuum. Both RT and identification data were collected. Based on earlier findings from the phoneme monitoring literature (Savin & Bever, 1970; McNeill & Lindig, 1973), subjects in the identification task were expected to respond faster when using word labels than when using phoneme labels. However, it was less certain what the shape of the identification functions for the two types of labels would be. One possibility is that due to the semantic content of words, listeners would in most normal situations have a tendency to focus more attention on words as wholistic units rather than on their constituent phonemes. To the extent that this is an accurate account of a listener's perceptual strategy, subjects would be expected to have more familiarity with words as perceptual categories than with phonemes as perceptual categories. Therefore, it might be expected that the slopes of the identification functions would differ depending upon whether word or phoneme labels were used to classify the speech stimuli. Specifically, the slope of the identification function for subjects using word labels was predicted to be steeper than the slope of the identification function for subjects using phoneme labels.

Method

Subjects. A total of 21 subjects were tested. Subjects received course credit in an introductory psychology course for their participation. All subjects were native speakers of English and reported no history of a speech or hearing disorder at the time of testing.

Stimuli. Two synthetic speech continua were constructed using the Klatt software synthesizer program (Klatt, 1978). One continuum had the words "BAD" and "PAD" as endpoints and the other had "RATE" and "LATE" as endpoints. In the BAD-PAD continuum, the critical parameter manipulated was voice onset time (VOT), which ranged from 10ms to 40ms. In the RATE-LATE continuum, the critical parameter manipulated was the initial frequency of the third formant, which varied from 1880 Hz to 2600 Hz. Each continuum contained seven synthetic stimuli. Spectrograms of endpoint stimuli for each continuum are shown in Figure 1.

Insert Figure 1 about here

Procedure. Subjects were tested in a quiet room. Each subject was seated at a booth in which a CRT monitor was located at eye level. Stimuli were presented over matched and calibrated TDH-39 headphones at 80 dB SPL as measured by a Hewlett Packard VTVM. The sequence of events in each trial began with a prompt presented on the monitor indicating the trial was about to begin. Then, two labels were presented on the lower half of the monitor, one in each corner. Presentation of labels occured under two conditions: In one condition, the two labels corresponded to the minimal pairs presented over the headphones, RATE-LATE or BAD-PAD. That is, the orthographic equivalants of the stimuli presented auditorily were presented on the monitor. This condition was called the "word labelling" condition. In the other condition, the two labels corresponded to the initial phoneme of the stimuli presented over the headphones, R-L or B-P. This second condition was called the "phoneme labelling" condition. The labels appeared on the monitor 500ms before the auditory stimulus. The labels remained on the monitor during presentation of the auditory stimulus and until a response by the last subject in the group was collected. Thus, the duration of the response labels on the monitor varied from trial to trial depending on how quickly subjects responded.

Subjects were told that their task was to identify what they heard over their headphones as quickly and as accurately as possible using the labels that appeared on the CRT monitor. Subjects in the word labelling condition were told that they would be listening to words whereas subjects in the phoneme labelling condition were told that they would be listening to speech sounds. Subjects in each condition were given essentially the same instructions with the exception of what would appear on the monitor.

Subjects were presented with a practice block of eight trials in order to familiarize them with the task. In the actual experiment itself, each stimulus item was presented 20 times, resulting in a total of 280 trials. Stimuli and labels were presented in a random order. The location of the labels on the CRT screen was consistent within an experimental session (e.g., BAD was always presented on the right and PAD was always preesnted on the left) but counterbalanced among different sessions. Although intertrial intervals varied somewhat according to subject's latencies, each trial took approximately 4s to complete. Altogether, each experimental session took approximately 35 minutes to complete.

Results and Discussion

Subject's responses were tabulated using two measures, the RT to each stimulus and the label used to identify each stimulus. The RT data will be described first. Mean latencies were calculated for each stimulus item in each continuum for both labelling conditions. These data are shown in Figure 2. The top panel shows the data for the RATE-LATE continuum while the bottom panel shows the data for the BAD-PAD continuum. For both, continua the response latencies for identifying stimuli in the word labelling condition were consistently faster than the latencies for identifying stimuli in the phoneme labelling condition. Also, latencies were longer for stimuli from the midpoints of the continua than for endpoint stimuli. An analysis of variance with factors of label (word and phoneme), continuum (RATE-LATE and BAD-PAD),

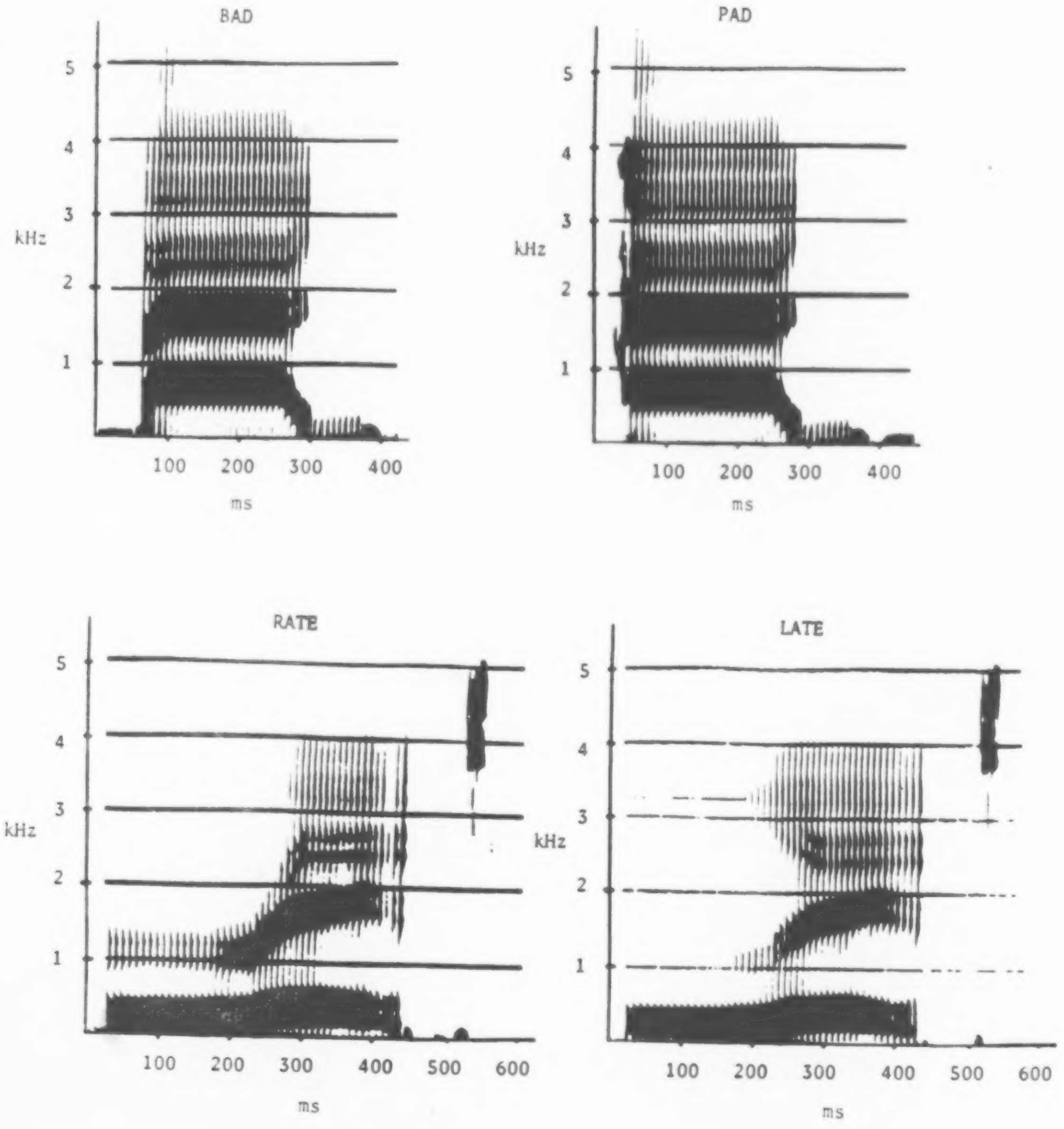


Figure 1. Spectrograms of the stimuli used in Experiment 1.

and position (items 1-7 in the continua) was used to assess the reliability of these observations. The ANOVA revealed a significant main effect for type of label [F(1,19)=7.92, p<.05], indicating that the time required for subjects to classify the stimuli depended on whether they were using word or phoneme labels. From an inspection of Figure 2, it is clear that subjects classified the stimuli faster when they used word labels than when they used phoneme labels. A significant main effect was also obtained for the position variable [F(6,114)=10.36, p<.001], indicating that the mean RT across the seven stimuli within each continua differed. There was no main effect of continua. However, a significant interaction between position and continuum was obtained [F(6, 114)=9.50, p<.001], indicating that there were differences in the latencies for identifying corresponding stimuli in each of the two continua. As noted above, subjects tended to have longer latencies for identifying stimuli from the middle of the continua than for identifying midpoint stimuli and that this tendency was greater for stimuli from the RATE-LATE continuum than for stimuli from the BAD-PAD continuum.

Insert Figure 2 about here

Identification data were also tabulated to show the percentage of different labelling responses that each stimulus received. These data are shown in Figure 3. Consistent labelling responses were obtained in each condition for each continua, regardless of labelling condition. In each case, the identification function showed a sharp transition as one label replaces another at some midpoint in the continuum. Since subjects reliably classified the stimuli into distinct categories based on the labels provided, the latency data was validated: a systematic relationship existed between categorization responses and response latencies. In order to test the predictions made regarding the slopes of the identification functions for the word and phoneme labelling conditions, the labelling data were fitted to a cummulative normal distribution. For the BAD-PAD continuum, the slope of the ID function was greater in the word labelling condition than in the phoneme labelling condition (*t* = 2.08, p<0.05). However, there was no reliable difference between the slopes of the ID functions in the word and phoneme labelling conditions for the RATE-LATE continuum (*t* = 0.24, p>0.05). Thus, the predictions made about the slopes of the ID functions for the two labelling conditions were only partially upheld for one stimulus continuum.

Insert Figure 3 about here

To summarize the results of the first experiment, when subjects were required to classify word stimuli differing only in initial phonemes, an increase in response time was obtained when subjects used phoneme labels compared to word labels. If viewed in one context, this finding is surprising; to identify the stimulus, subjects should need to hear only the first phoneme since it is only this phoneme that differentiates one category from the other. Furthermore, since the labels are identical with reference to the first phoneme of the stimulus, subjects might be expected to attend to only the first phoneme of the stimulus they hear and to only the first letter of the label they see on the monitor. Such a strategy would seem to be the

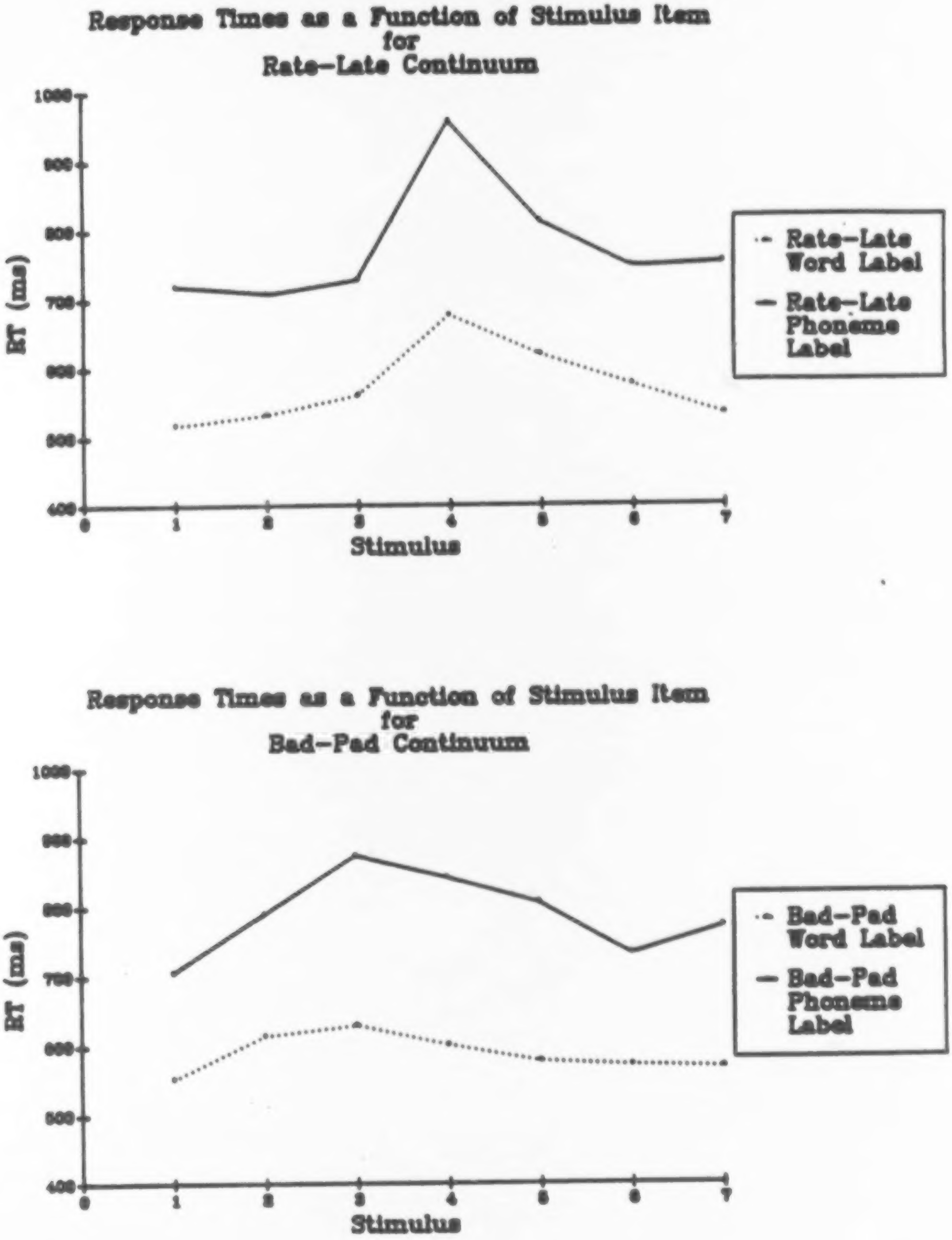


Figure 2. Mean latencies for identifying each stimulus item using word and phoneme labels. The top panel show latencies from the RATE-LATE continuum and the bottom panel show latencies from the BAD-PAD continuum.

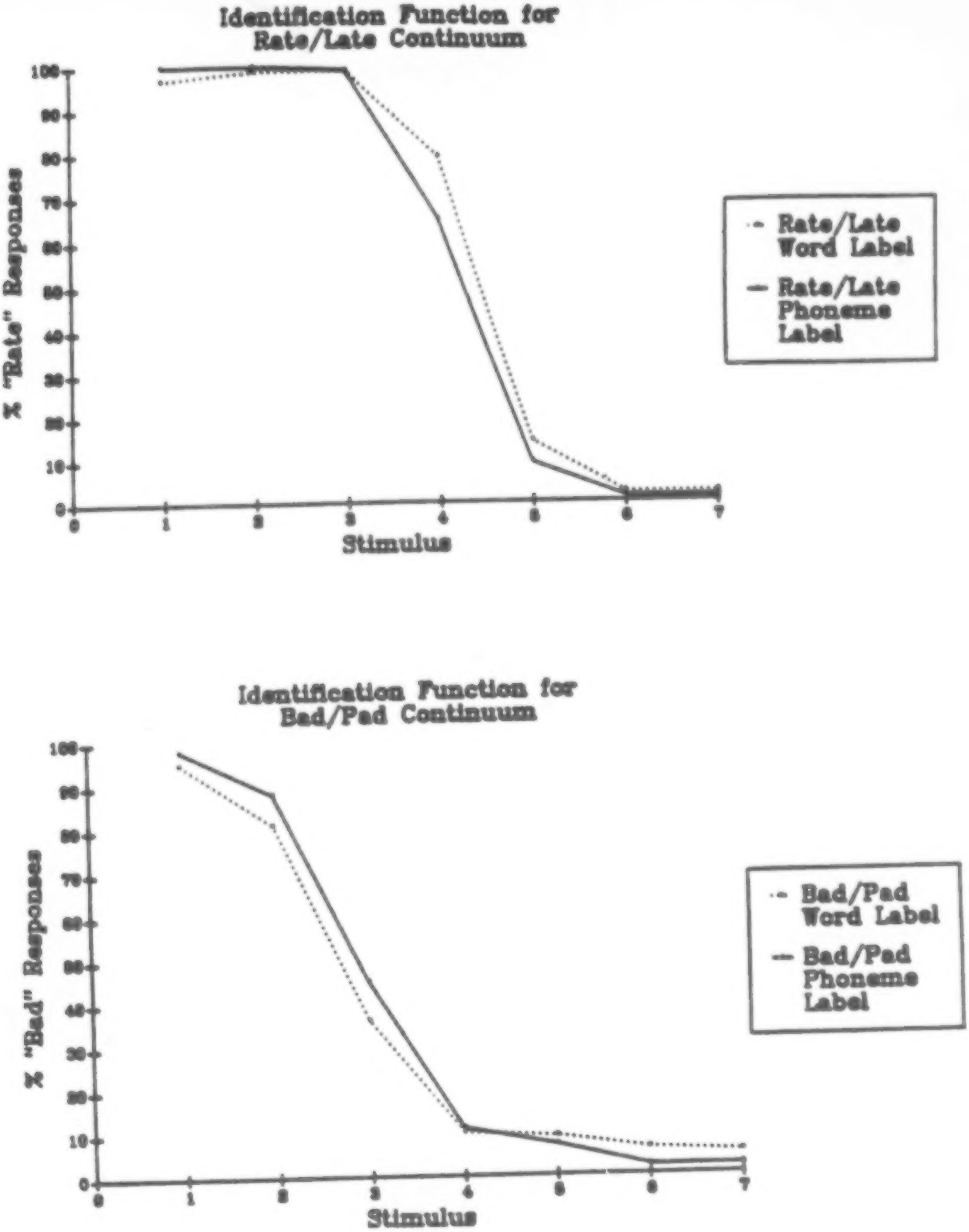


Figure 3. Identification functions using word and phoneme labels for the RATE-LATE continuum (top panel) and the BAD-PAD continuum (bottom panel).

simplest to successfully meet the demands of the task and would likely result in no difference in the latencies for word and phoneme labels. However, this did not happen.

A possible explanation for the pattern of results is that the perceptual and cognitive processing of a word may be intrinsically more simple and therefore faster than making judgements about the components that comprise a word. If in most tasks individuals use words as the primary unit of perception and production due to their meaningfulness, it might be expected that phonemes, which have no intrinsic meaning, would be less familiar to listeners than words. Furthermore, it could also be the case that the primary means by which information in the lexicon is accessed is by using the word as a holistic unit. This view is compatible with Klatt's (1979) LAFS model, in which direct lexical access to word knowledge is accomplished through spectral information mapped on to a level of representation corresponding to words. Further information about the phonetic structure of the word, such as the initial phoneme, would either be computed or retrieved subsequent to the initial lexical look-up. From this perspective, to extract information about what phonemes comprise a word would take extra time due to the extra processing beyond that required to decide that a particular word had been presented. However, given the data from phoneme monitoring experiments (eg., Cutler et al., 1987), it is not likely that top-down phonological information is the sole means by which subjects can make decisions about the phonemic composition of words.

A related explanation for the pattern of results observed in the first experiment concerns the sequencing of events in the experiment and how this may have affected subjects' performance. On each trial, presentation of the label which the subject used to classify the stimulus preceded presentation of the stimulus by a short period of time. The presentation of the label likely created some expectancy about what stimulus would be presented and also how that stimulus would be processed. In other words, if word labels were presented, the corresponding word units in the subject's lexicon may have been activated, thus facilitating response time when the actual stimulus was presented a short time later. If only the initial letters corresponding to the first phoneme in the stimulus were presented, no such facilitation would have occured, since the letters alone would be insufficient to activate an appropriate lexical representation. Subjects would simply take longer to classify the stimuli using phoneme labels than using word labels due to insufficient priming by the phoneme labels as compared to the word labels.

In light of these results and our speculations about their underlying basis, it seemed reasonable to carry out a second experiment in which subjects were asked to classify nonwords. By their very nature nonwords do not have the same lexical status as words and therefore would not be expected to exert the same influence on processing as would words. Also, to make sure that the results obtained in the first experiment were reliable, the entire experiment was replicated. Therefore, the second experiment had two goals: first, to see what results would be obtained when subjects were required to classify nonwords using the same experimental procedure as used in the first experiment and second, to replicate the results obtained in the first experiment.

Experiment 2

The design of the second experiment was modelled closely after the first experiment. With the exception of an added condition in which a set of nonword synthetic stimuli were used, the two experiments were essentially identical. Subjects used both word and phoneme labels to classify speech stimuli into different categories. Results from the nonword condition should provide insights into the mechanisms reponsible for the results observed when subjects classified word stimuli. Several predictions about the outcome of the second experiment were made. First, for the word stimuli, results similar to those obtained in the first experiment would also be found in the second experiment. For the RT data, replication of the first experiment was considered to be quite certain since the RT differences for the two labelling conditions were reliable. With respect to the ID data, replication was less certain since no effect of slope was found for one of the continua in Experiment 1. It was less obvious in the nonword condition what the pattern of results would be. A priori, there was reason to believe that the RTs in the phoneme labelling condition would be slower than found in Experiment 1 since Rubin et al. (1976) found that RTs for detecting initial phonemes in words was faster than for detecting initial phonemes in nonwords. Also, RTs for the nonword stimuli using pseudoword labels were predicted to be slower than those obtained in Experiment 1 due to a lack of top-down, lexically-based faciltiation. In predicting how the RTs for the nonword stimuli in the pseudoword and phoneme labelling condition would compare with each other, it was not known if there would be any difference between the RTs or if one type of label would result in faster responding than the other.

Method

Subjects. A total of 49 subjects participated in the second experiment. All were native speakers of English and reported no history of a speech or hearing disorder. Subjects received class credit for their participation.

Stimuli. The stimuli in the word condition were identical to those used in Experiment 1 with one exception. One minor modification was made to the stimuli from the RATE-LATE continuum: the first one hundred ms was deleted from each stimulus in this continuum because the /r/ - /l/ portion of the stimuli sounded too long. After deletion of this initial portion, the simuli sounded more natural and the essential quality of "r" or "l" was preserved. No change was made to the stimuli from the BAD-PAD continuum.

Nonword stimuli were modelled after the word stimuli. The nonword stimuli were identical to the word stimuli except for the final consonant which was changed to produce a nonword. The phoneme /t/ in the RATE-LATE stimuli was replaced with the phoneme /b/, while the phoneme /d/ in the BAD-PAD stimuli was replaced with the phoneme /v/. Thus, RABE-LABE and BAV-PAV nonword continua were generated, each containing seven stimuli.

Procedure. Experimental procedures in the second experiment were similar to those in the first experiment. As in Experiment 1, subjects who listened to word stimuli were tested in two separate groups, one given word labels and the other given phoneme labels with which to identify the stimuli. Subjects presented with word stimuli were given the same instructions as subjects in the first experiment. Subjects in the nonword condition followed the same procedure with one exception. Those subjects who listened to nonword stimuli who were in the word labelling condition were presented with the orthographic equivalents of the nonword stimuli, RABE-LABE or BAV-PAV. They were told they would be listening to pseudowords which they would be asked to identify using the labels appearing on the CRT monitor in front of them. For simplicity, I will refer to this condition as a "word labelling" condition while acknowledging the inexactitude of this nomenclature. Subjects in the nonword condition who were asked to use phoneme labels to identify the stimuli were given the same instructions and were given the same labels as those subjects in the comparable word stimuli condition.

Results and Discussion

As in Experiment 1, subjects' responses were tabulated using two dependent measures, the mean RT for each stimulus and the response label assigned to each stimulus. The RT data will be presented first. Mean RT values for the identification of word and nonword stimuli using word and nonword labels are shown in Figure 4. For word stimuli, RTs were faster for subjects who used word labels than for subjects who used phoneme labels. For nonword stimuli, RTs were faster for subjects who used phoneme labels than for subjects who used word labels. Thus, the effect of using different labels was replicated for the word stimuli.

Insert Figure 4 about here

Figures 5 and 6 show the mean latencies for individual stimuli within each continuum for both stimulus conditions and for both labelling conditions. Mean latencies for stimuli beginning with /r-l/ are shown in Figure 5 while Figure 6 shows the mean latencies for stimuli beginning with /b-p/. The top panel in each figure shows latencies for the word stimuli while the the bottom panel shows latencies for the nonword stimuli. Subjects identified all of the word stimuli faster when using word labels than when using phoneme labels whereas subjects identified all of the nonword stimuli faster when using phoneme labels than when when using word labels. The effect is consistent for both sets of stimuli. Some differences among continua are also apparent when examining these two figures. Latencies for stimuli from the /r-l/ continua display a peak near the midpoints of the continuum for the word stimuli while no such peak is found for stimuli from the nonword condition. However, latencies for stimuli from the /b-p/ continua do not display as large a peak in either the word or nonword conditions as the latencies for stimuli from the /r-l/ continua. This pattern is similar to what was found in Experiment 1 for word stimuli. Latencies were longer for stimuli from the midpoint region than for stimuli from the endpoints of the RATE-LATE continuum but this was not the case for stimuli from the BAD-PAD continuum in which a much smaller peak was observed.

Insert Figures 5 and 6 about here

An ANOVA with factors of stimulus (word and nonword), label (word and phoneme), continuum (/r-l/ and /b-p/), and position (items 1-7 in the continua) was used to assess the reliability of the observations noted above. No significant main effects for either stimulus or label were obtained. There was a significant main effect for continuum [$F(1,45)=64.84$, $p<0.001$] due to the overall faster RT for stimuli from the /r-l/ continua than for stimuli from the /b-p/ continua. There was also a significant main effect for position [$F(6, 270)=5.99$, $p<0.001$] due to the tendency for stimuli from the endpoints of the continua to be identified more quickly than stimuli from the middle of the continua. Several significant interactions were observed. First, there was a significant interaction between stimulus and label

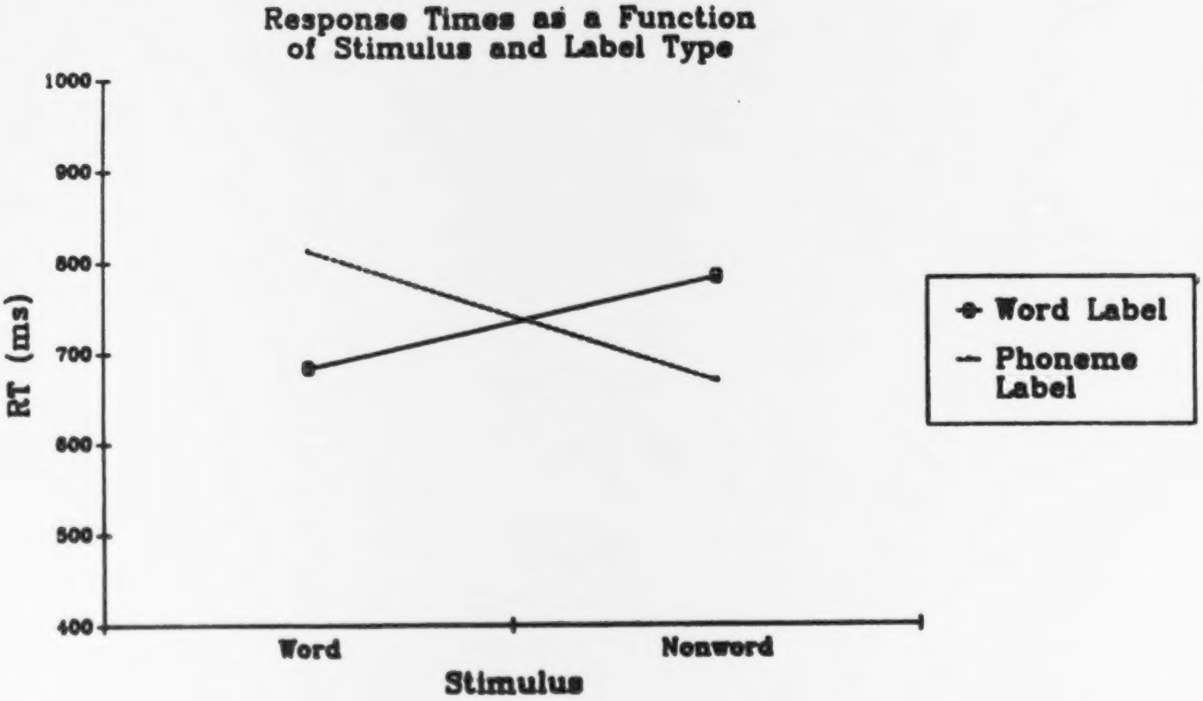
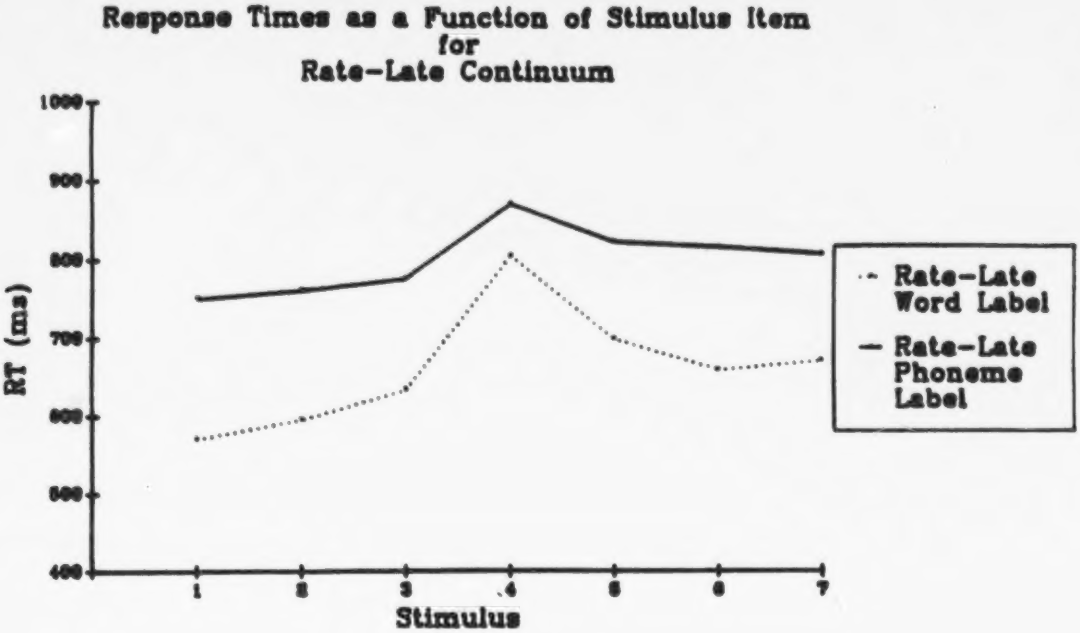
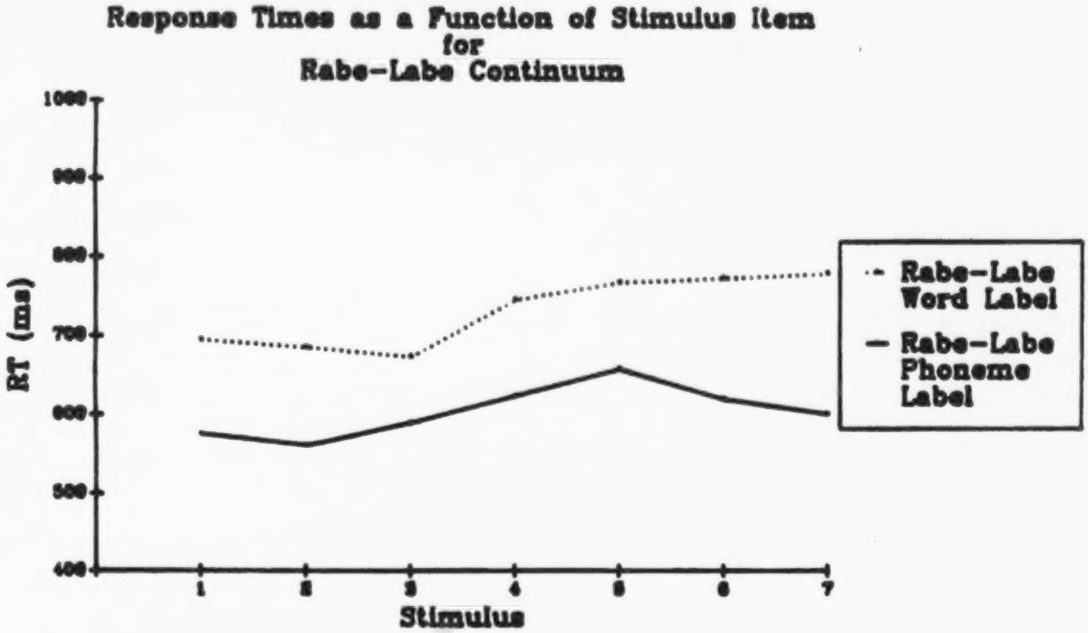


Figure 4. Mean latencies for identifying the word and nonword stimuli using word and phoneme labels.



Response Times as a Function of Stimulus Item for Rate-Late Continuum



Response Times as a Function of Stimulus Item for Rabe-Labe Continuum

Figure 5. Mean latencies for identifying each word stimulus item using word and phoneme labels. The top panel shows latencies for stimuli from the RATE-LATE continuum and the bottom panel shows latencies for stimuli from the RABE-LABE continuum.

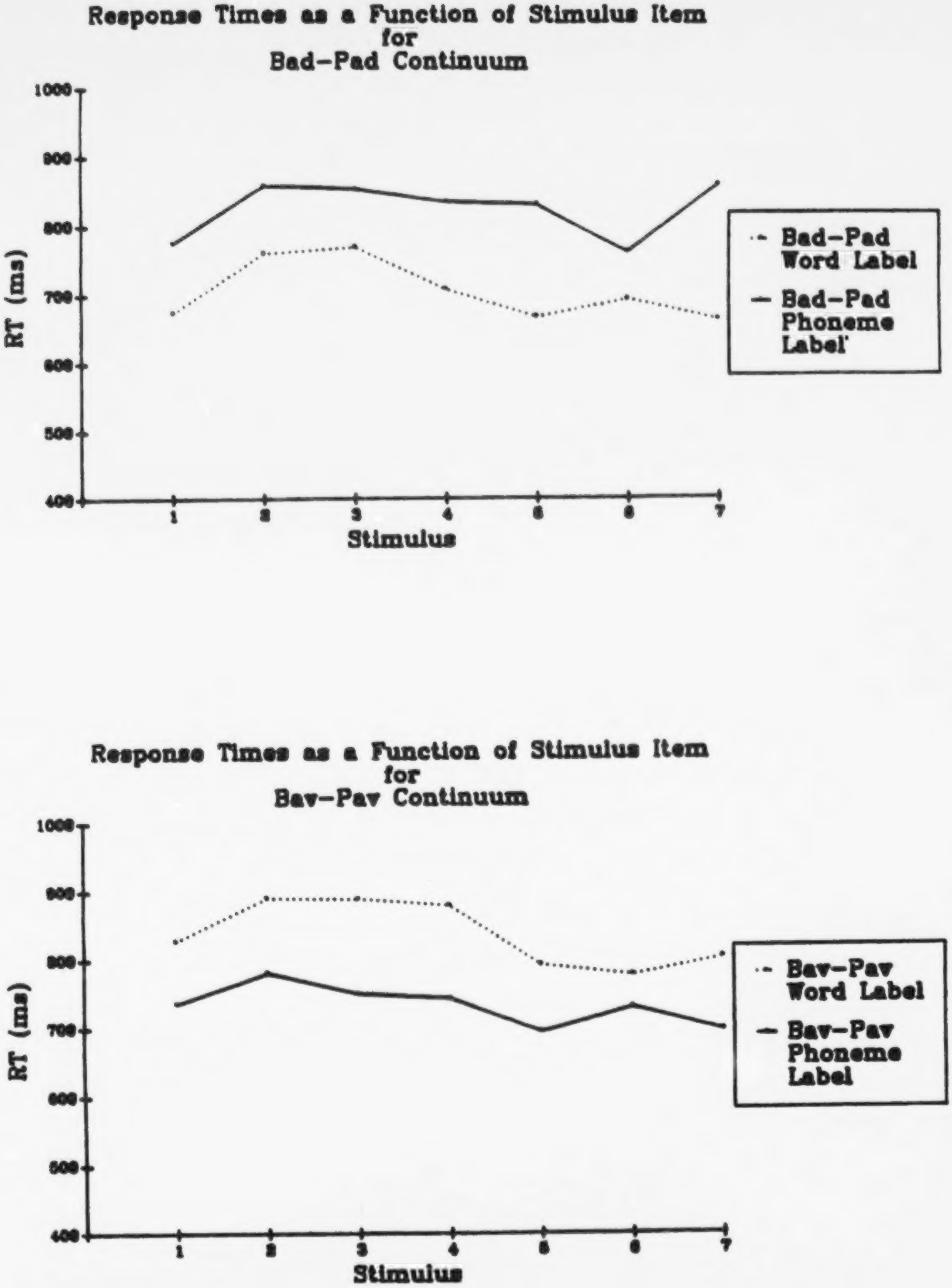


Figure 6. Mean latencies for identifying each nonword stimulus item using word and phoneme labels. The top panel shows latencies for stimuli from the BAD-PAD continuum and the bottom panel shows latencies for stimuli from the BAV-PAV continuum.

[F(1,45)=7.78, p<0.01], indicating that the word stimuli were identified faster when listeners used word labels than when they used phoneme labels whereas the nonword stimuli were identified faster when listeners used phoneme labels than when they used word labels. Second, there was a significant interaction between continuum and stimulus [F(1,45)=20.53, p<0.001]. The interaction between stimulus and continuum is shown in Figure 7. The difference in mean RT between the two continua, /r-l/ and /b-p/, is greater for the nonword stimuli than for the word stimuli. At the present time, the source of this interaction is not known. A third significant interaction was found between continua and position [F(6, 270)=12.13, p<0.001] due to a difference between the two continua in which stimulus items were associated with the slowest RTs. For the /r-l/ continua the slowest RTs were for stimuli from the middle of the continua whereas for the /b-p/ continua the slowest RTs were for stimuli from near the /b/ endpoint.

Insert Figure 7 about here

The identification functions for the /r-l/ and /b-p/ continua are shown in Figures 8 and 9, These data were tabulated from the proportion of times each label was assigned to each stimulus item. In each case, the ID functions indicate that subjects classified the stimuli from the continua into two categories regardless of the label used or the lexicality of the items. Overall, the ID functions are essentially the same as those obtained in Experiment 1. In order to test the prediction that the slope of the functions would differ depending upon the label and the stimulus, the ID data were fitted to a cumulative normal function and slopes were computed. The slope data were calculated individually for each subject. An ANOVA with factors of label (word and phoneme), stimulus (word and nonword), and continuum (/r-l/ and /b-p/) was used to determine if significant differences existed between the slopes in the different conditions. No main effects were obtained. However, there was a significant three-way interaction among the three variables (label, stimulus, and continuum) [F(1, 45)=5.19, p<0.05]. Figure 10 shows the mean slopes for the word and nonword stimuli using the word and phoneme labels for both the /r-l/ and /b-p/ continua. For the /r-l/ continua, the slope of the labelling function for words is steeper when word labels are used to classify the stimuli but for monwords, the slope of the labelling function is steeper when phoneme labels are used. For the /b-p/ continua, the slope of the labelling function for both words and nonwords is steeper when word labels are used than when phoneme labels are used. Thus, the three-way interaction appears to be due primarily to the lack of a crossover for the slope values when using word and phoneme labels to classify the word and nonword /b-p/ stimuli compared to the effect found with the /r-l/ stimuli.

Insert Figures 8 and 9 about here

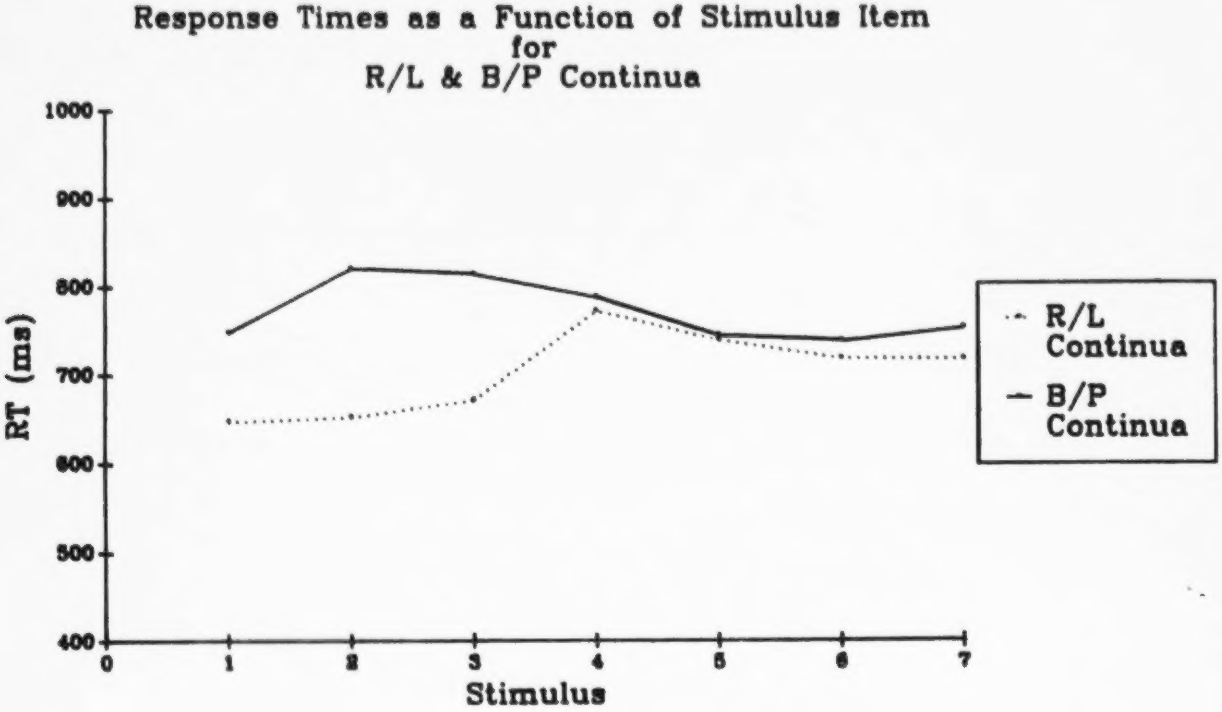


Figure 7. Mean latencies for identifying each stimulus item from the /r-l/ and /b-p/ continua.

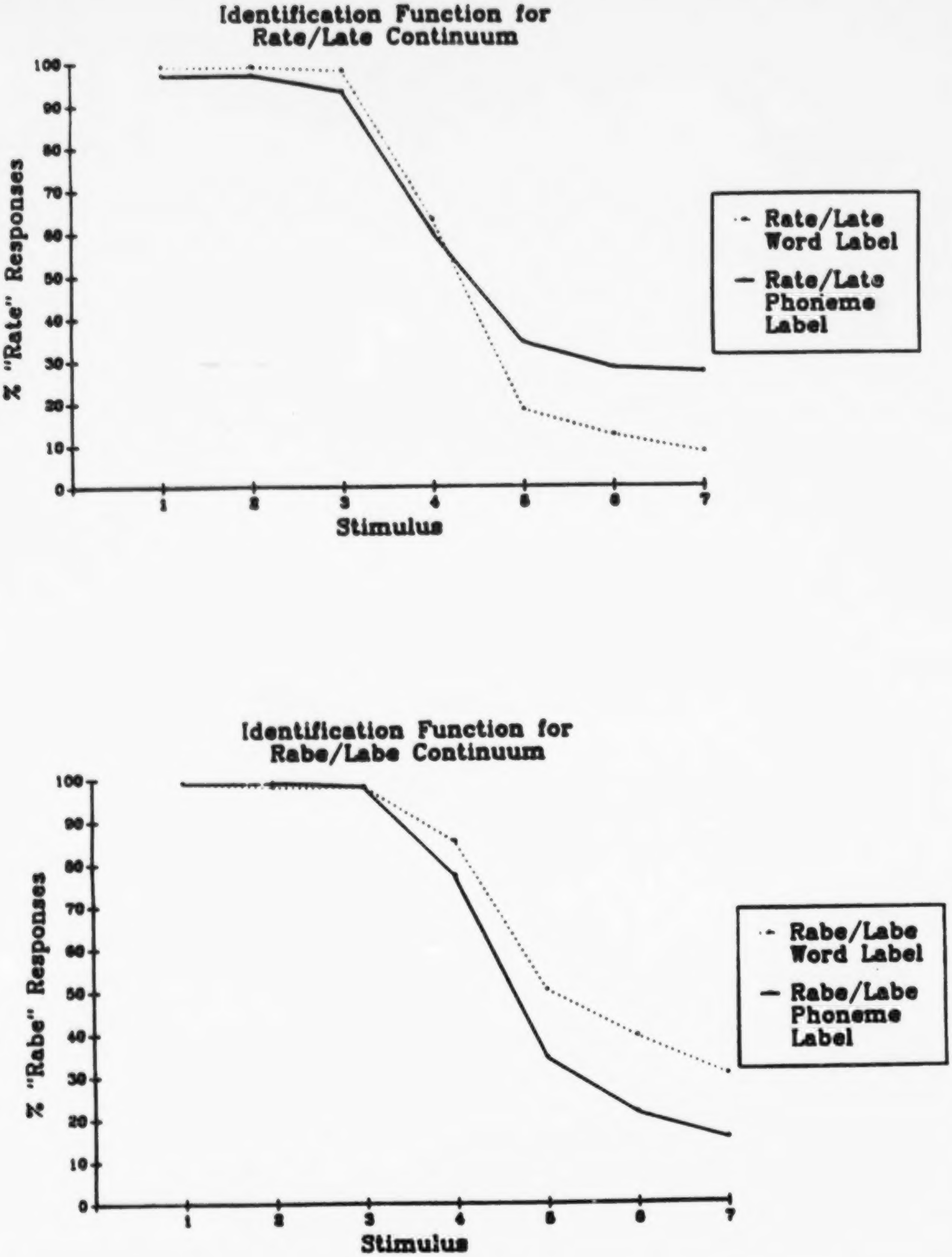


Figure 8. Identification functions using word and phoneme labels for the RATE-LATE continuum (top panel) and the RABE-LABE continuum (bottom panel).

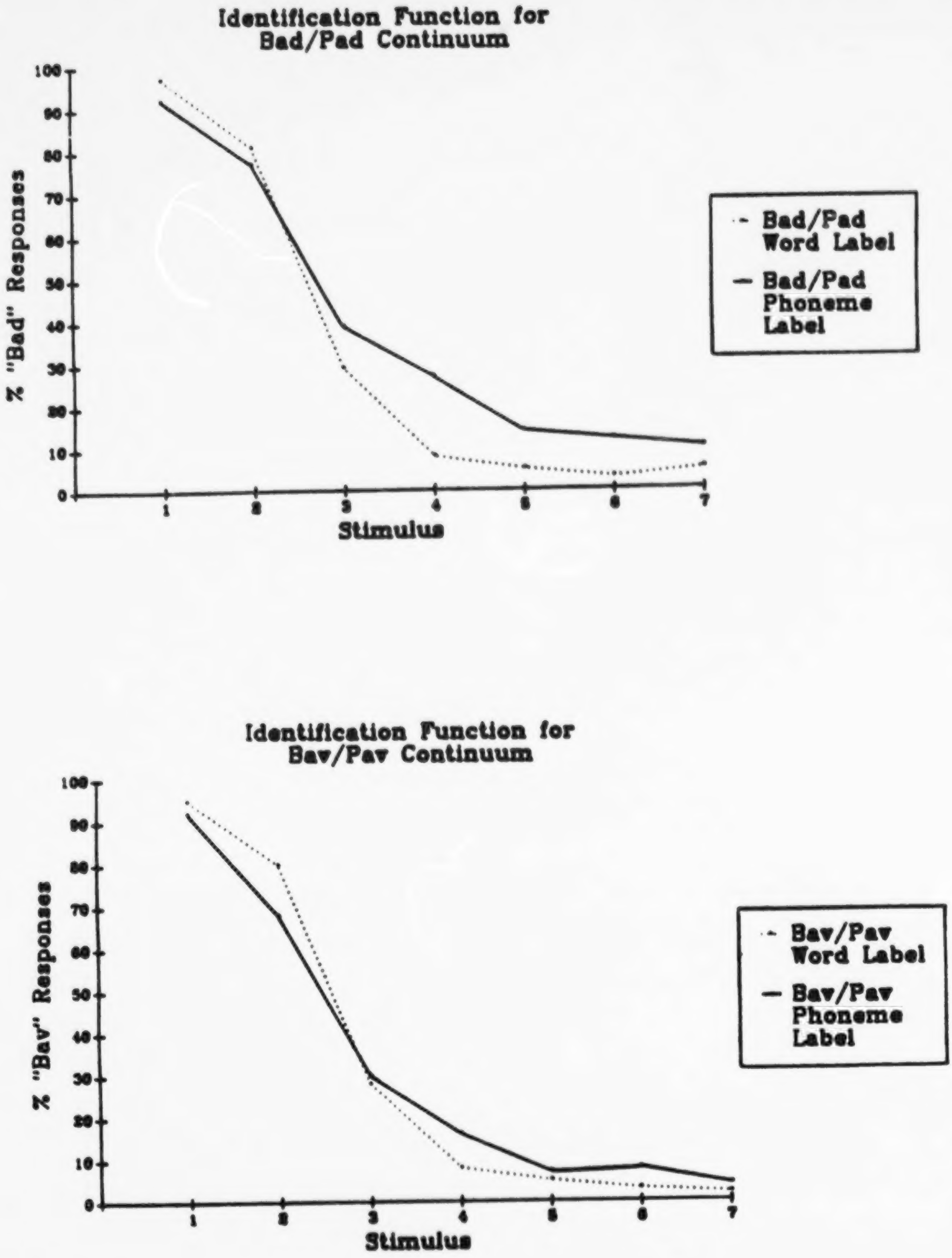


Figure 9. Identification functions using word and phoneme labels for the BAD-PAD continuum (top panel) and the BAV-PAV continuum (bottom panel).

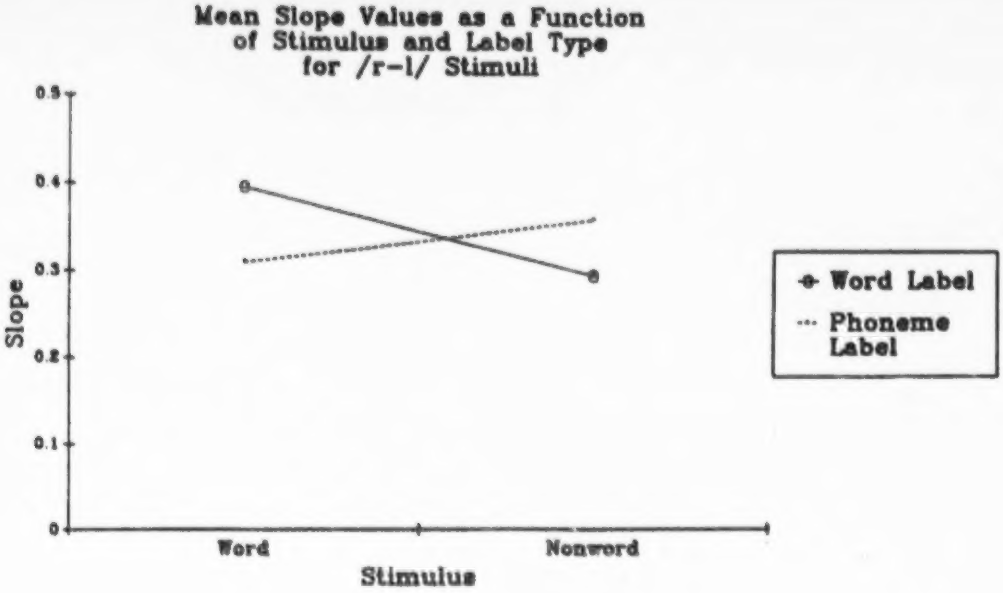
Insert Figure 10 about here

In general, the ID data provide converging evidence for the pattern of results observed with the RT data. Subjects responded more quickly and with more sharply defined category boundaries when they used word labels to classify word stimuli than when they used phoneme labels. In contrast, subjects tended to respond more quickly and with more sharply-defined category boundaries when they used phoneme labels to classify nonword stimuli. The results for the /r-l/ continua are somewhat more consistent than those obtained using the /b-p/ continua. In the case of the /r-l/ stimuli, there is a clear correspondence between the results of the RT and ID measures. For the /b-p/ stimuli, the ID data match the RT data except for the nonword stimuli. The overall pattern of data suggests that different mechanisms are responsible for the identification of word and nonword stimuli and that using word and phoneme labels maybe a useful way to dissociate these two mechanisms.

General Discussion

The results of Experiment 1 indicated that using word and phoneme labels to classify speech produced different effects. Subjects were faster to identify spoken words when they used a word label than when they used a label corresponding to the initial phoneme. Converging evidence for the efficacy of word labels over phoneme labels for identifying spoken words was also obtained in the analyses of the identification functions. For one of the two continua, subjects demonstrated a more sharply-defined category boundary when the word label was used. Taken, together, these results suggest that the presentation of a word label shortly before the presentation of the speech stimulus caused the activation of the lexical entries corresponding to the labels. Activation of the lexical entries likely served to make available phonological information associated with the activated entries. When the speech stimulus was presented, the phonological information from the activated lexical entries facilitated decisions about which stimulus was actually heard compared to the condition in which phoneme labels were used. It is unlikely that presentation of the phoneme labels activated any lexical information so that decisions about what spoken word had been presented in that condition presumably proceeded from a bottom-up prelexical analysis of the signal. It is also possible that the identification of the word stimuli using phoneme labels used phonological information which, because of the lack of prior lexical activation from the phoneme labels, was slowed compared to the RTs in the word labelling condition.

The results of Experiment 2 indicated that the findings obtained in the first experiment were reliable. Moreover, the results obtained with the nonword stimuli showed a pattern of responding opposite to that obtained with the word stimuli. Subjects were faster using phoneme labels to classify nonword stimuli than when using pseudoword labels. As in the first experiment, some converging evidence for the pattern of results found using the RT measure was also found in the identification data. For both continua, subjects showed more sharply defined category boundaries when classifying the word stimuli using word labels. However, for one of the continua, subjects showed a more sharply defined category boundary when classifying the nonword stimuli using phoneme labels.



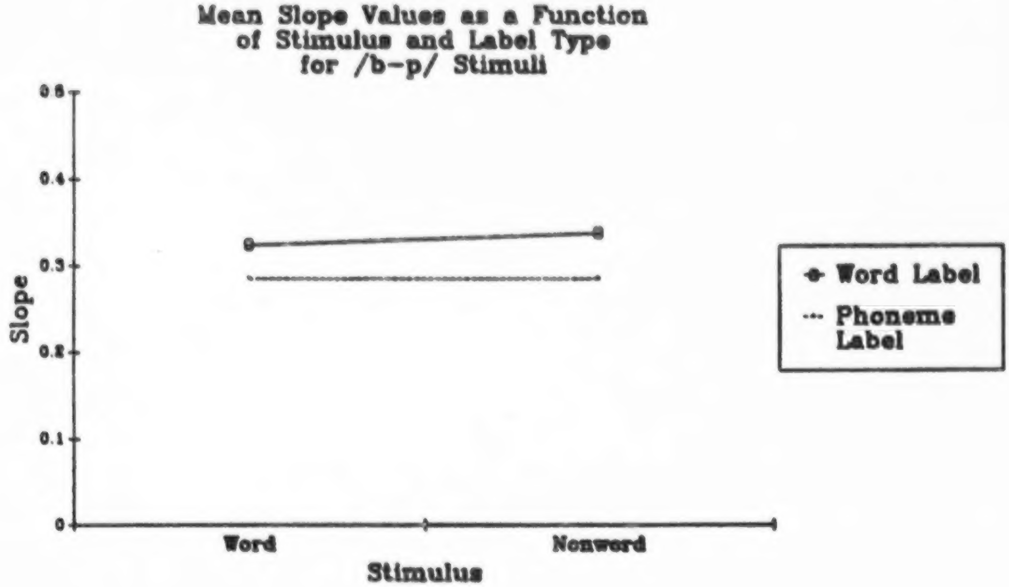


Figure 10. Mean slope values of the identification functions using word and phoneme labels for the /r-l/ and /b-p/ word and nonword stimuli.

The results obtained in Experiment 2 clearly demonstrated the role played by the lexical status of the stimulus and label. Presentation of a pseudoword label to subjects who were listening to nonword stimuli did not facilitate response times for identification of the nonword stimuli. In fact, the possibility exists that the pseudoword labels may have actually inhibited RTs due to a search of the lexicon for the nonexistant pseudowords. In contrast, the use of phoneme labels to identify the nonword stimuli may have caused subjects to focus their attention at a prelexical level, facilitating the use of bottom-up information in making the identification response. The reasoning behind this explanantion for the results of Experiment 2 is that it would have been impossible in the nonword condition for any kind of postlexical phonological information to be responsible for the identification of the stimuli since by definition the nonword stimuli are not present in the listener's lexicon and therefore no phonological information exists for these stimuli. Consequently, decisions about what stimulus was presented in the nonword condition must have been based on bottom-up pre-lexical information. For those subjects who were presented the word stimuli, the possiblity exists that both pre- and postlexical information could have been used to make the identification response. However, since the RTs for word labels were faster than for the phoneme labels, the explanantion for the results in this case favors a lexically-based, phonologically-derived response.

The explanation of the results obtained in Experiment 1 and 2 is tentative at this ponit. But, the account of the mechanisms responsible for the results of the present experiment is consistent with the explanations offered by others for results from phoneme monitoring experiments (Cutler et al., 1987) and identification experiments (Ganong, 1980; Fox, 1984; and Connine & Clifton, 1987). Lexical information can be used to focus of attention at a lexical level in which phonological information becomes activated. However, under some circumstances, such as when the stimuli are nonwords, attention can be focused at a prelexical level in order to make identification decisions. Unfortunately, given the limited scope of the present experiment, the results do not permit the selection of one model of speech perception over another. The results of the present experiment are equivocal as to whether an interactive model such as TRACE (McClelland & Elman, 1986) or a parallel model such as the race model (Cutler & Norris, 1979) can best deal with the existing psychological data.

In order that the mechanisms responsible for the findings obtained in the present experiment may be more fully understood, several future studies are planned. First, it is important to understand the extent to which lexical status effects the usefulness of the label as either facilitating or inhibiting RT. One way to vary the lexical status of the stimuli and labels is to use words of varying frequencies. High frequency words are probably more word-like than low frequency words, especially words that occur only rarely. If a set of stimuli varying in frequency was used in a task such as the one used here, it is possible that as word frequency was decreased, the magnitude of the RT difference between word and phoneme labels would also decrease. As the frequency of the label corresponding to the word stimulus was decreased, the likelihood that a lexical entry would be activated would also decrease, therefore reducing the facilitory effect of the label on identification. A second study should investigate the temporal interval between the presentation of the label and the presentation of the speech stimulus and its effect on identification (cf. Neely, 1977). If this interval was made successively shorter, the effect of the word label on the activation of a lexical entry would probably be reduced, even to the point where no advantage for the word label would be obtained compared to using a phoneme label when listening to word stimuli. Manipulation of word frequency and the temporal interval between the label and speech stimulus would therefore provide information about how lexical information can be used to make judgements about the components that comprise a word and under what circumstances pre- or postlexical information is used to make such judgements.

In conclusion, the results of the present study provide several insights into the way listeners can make decisions about the internal composition of words. It appears that listeners use postlexical phonological information as well as prelexical phonetic information to determine the phonetic composition of speech stimuli. The use of the labelling paradigm in the present experiments offers an additional methodology to explore the mechanisms of speech perception that complements previously used techniques such as phoneme monitoring and word identification tasks.

References

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P., Kennedy, L., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General*, *117*, 21-33.

Connine, C., & Clifton, C. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 291-299.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, *19*, 141-177.

Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In (W. Cooper & E. Walker, Eds.) *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*, Hillsdale, NJ: Erlbaum.

Dell, G.S., & Newman, J.E. (1980). Detecting phonemes in fluent speech. *Journal of Verbal Learning and Verbal Behavior*, *19*, 608-623.

Foss, D., & Blank, M. (1980). Identifying the speech codes. *Cognitive Psychology*, *12*, 1-31.

Foss, D., & Gernsbacher, M. (1983). Cracking the dual code: Toward a unitary model of phoneme identification. *Journal of Verbal Learning and Verbal Behavior*, *22*, 609-632.

Foss, D., & Swinney, D. (1973). On the psychological reality of the phoneme: Perception, identification, and consciousness. *Journal of Verbal Learning and Verbal Behavior*, *12*, 246-257.

Fox, R. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 526-540.

Ganong, W. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 110-125.

Healy, A., & Cutting, J. (1976). Units of speech perception: Phoneme and syllable. *Journal of Verbal Learning and Verbal Behavior*, *15*, 73-83.

Klatt, D. (1979). Speech perception: A model of acoustic-phonetic and lexical access. *Journal of Phonetics*, *7*, 279-312.

Liberman, I., Shankweiler, D., Liberman, A., Fowler, C., & Fischer, F. (1977). Phonetic segmentation and recoding in the beginning reader. In (A. Reber & D. Scarborough, Eds.) *Toward a psychology of reading*. Hilldale, NJ: Erlbaum.

McClelland, J., & Elman, J. (1985). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1-86.

McNeill, D., & Lindig, K. (1973). The perceptual reality of phonemes, syllables, words, and sentences. *Journal of Verbal Learning and Verbal Behavior*, *12*, 419-430.

Morton, J., & Long, J. (1976). Effect of word transitional probability on phoneme identification. Journal of Verbal Learning and Verbal Behavior, 15, 43-51.

Neely, J. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited capacity attention. Journal of Experimental Psychology: General, 106, 226-254.

Peterson, G., & Lehsite, I. (1960). Duration of syllable nuclei in English. Journal of the Acoustical Society of America, 32, 693-703.

Pisoni, D., & Luce, P. (1987). Acoustic-phonetic representations in word recognition. Cognition, 25, 21-52.

Rubin, P., Turvey, M., & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. Perception & Psychophysics, 19, 394-398.

Samuel, A., & Ressler, W. (1986). Attention within auditory word perception: Insights from the phonemic restoration illusion. Journal of Experimental Psychology: Human Perception and Performance, 12, 70-79.

Savin, H., & Bever, T. (1970). The nonperceptual reality of the phoneme. Journal of Verbal Learning and Verbal Behavior, 9, 295-302.

[RESEARCH ON SPEECH PERCEPTION Progress Report No. 13 (1987) Indiana University]

Talker Variability and the Recall of Spoken Word Lists: A Replication and Extension

John Logan and David B. Pisoni

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, IN 47405

*This research was supported, in part, by NIH Research Grant NS-12179-11 to Indiana University in Bloomington.

Abstract

Martin, Mullennix, Pisoni, and Summers (1987) have recently examined recall of lists of words spoken by single and by multiple talkers. Using a serial recall task, they found that for early list items, recall was better for lists produced by single talkers than for lists produced by multiple talkers. The present paper reports two preliminary experiments that were designed to follow up on the findings of Martin et al. Experiment 1 examined differences in recall between lists of words produced by single and multiple talkers using lists of equivalent length in both free and serial recall tasks. The confusability of the words was also manipulated by varying word frequency and phonetic similarity. In the serial recall task, we found that recall was better for early list positions for single talker lists whereas recall was better in late list positions for multiple talker lists. In the free recall task, results were more variable due to large individual differences in recall strategies although for early list positions, recall was significantly better for single talker lists. Highly confusable words were not recalled as accurately as low-confusability words, especially in early list positions. Experiment 2 used a serial recall task in which the order of the talker in the multiple talker lists was held constant from list to list. This manipulation was designed to determine if recall for items from the early part of the lists would be improved compared to Experiment 1. Results revealed increases in recall for early list items for the multiple talker lists. Subjects apparently use the voice cues from individual talkers to encode both item and order information in this task. Taken together, the present results demonstrate that talker variability affects encoding and/or rehearsal processes differently for early and late portions of the serial position curve. The results have implications for current conceptions of short- and long-term memory.

Talker Variability and the Recall of Spoken Word Lists: A Replication and Extension

Recently, Mullennix, Pisoni, and Martin (1987) carried out a series of experiments in which they examined the perception of monosyllabic words produced by a single talker compared to the same stimuli produced by several different talkers. Using several different tasks, they found that items produced by multiple talkers resulted in performance decrements compared to the same items produced by only one talker. Listener's accuracy and speed of response in both perceptual identification and naming tasks were adversely affected when the stimuli were produced by multiple talkers. These results suggested the operation of some form of perceptual normalization process that compensates for the different physical attributes of the different voices. A normalization process for different talkers might require some additional amount of time and processing capacity thus affecting those cognitive processes subsequent to normalization, such as word recognition and lexical access.

Based on the perceptual findings obtained by Mullennix et al. (1987), Martin, Mullennix, Pisoni, and Summers (1987) proposed that if normalization processes used when listening to different talkers affected the speed and accuracy of later cognitive processes, then these effects should be evident in other kinds of tasks as well. Specifically, Martin et al. examined listener's recall of lists of words spoken by single and multiple talkers. If the effects of talker normalization affect further levels of cognitive processing as shown by Mullennix et al. (1987), then presumably the effects should also be evident in memory-related tasks as well. Some earlier research on the effects of talker variability had shown that facilatory effects due to multiple talkers could be obtained in certain kinds of memory tasks. For example, Craik and Kirsner (1974) found that talker-specific features could be used to facilitate recognition memory for words if the same voice that was used to cue subjects was also used originally to present the words. On the other hand, using a serial recall task, Mattingly, Studdert-Kennedy, and Magen (1983) found that recall of early list items produced by multiple talkers was lower than recall of the same items produced by a single talker. In short, these early studies showed that using multiple talkers to produce a list of words can affect a listener's memory for those words, and that the effect depends on the task used and the position of the items within the list.

As a starting point for their investigation of the effect of talker variability on recall, Martin et al. reviewed the results of an earlier experiment carried out by Luce, Feustel, and Pisoni (1983) which was designed to examine differences in recall between natural and synthetic speech. Luce et al. found that for speech produced by a high-intelligibility synthesizer -- the MITalk system, performance on early list items was reduced compared to recall of natural speech. Luce et al. argued that the lower performance on early list items for the synthetic speech may have been due to increased encoding demands that affected the rehearsal of early list items. Because of the degraded acoustic-phonetic information in the synthetic speech, more effort and processing capacity was required to encode the sensory information into a phonetic representation compared to the effort required for natural speech. If the short-term memory (STM) system has a limited capacity as suggested by some theorists (e.g., Shiffrin, 1976), then the additional effort required to encode impoverished sensory information would reduce the amount of capacity available for subsequent rehearsal of list items and thus impair their transfer to long-term memory (LTM). Since it is commonly believed that the recall of early list items is due to retrieval of the items from LTM

(Atkinson & Shiffrin, 1968), any impairment of transfer of information to LTM would therefore result in reduced recall of early list items. Applying this reasoning to word lists produced by multiple talkers, the normalization process may require additional processing resources in working memory, thus affecting the time course and efficiency of processes used to map sensory input onto representations in LTM. Going one step further, it would seem reasonable to assume that the processing demands made by lists of words produced by multiple talkers might affect recall of items from early list positions in a manner similar to that found with synthetic speech.

Martin et al. carried out four experiments that examined listeners ability to recall lists of monosyllabic words produced by single and multiple talkers. In their first experiment, they examined serial recall for ten-word lists. They found that recall of early list items in the multiple talker condition was reduced compared to recall of items in the single talker condition. This result was consistent with the limited capacity STM argument described above: the normalization process required to encode speech produced by multiple talkers appears to affect the processes involved in the transfer of item and order information to LTM.

In a second experiment, Martin et al. examined free recall for lists of words containing 20 items produced by single and multiple talkers. A free recall task was used to explore the nature of the decrement in performance observed for early list items produced by multiple talkers in the first experiment. Compared to serial recall, the free recall task should make fewer demands on STM processes since encoding information about the order of items within each list is not required. Therefore, more processing resources in STM could be devoted to the transfer of items to LTM. In addition to the use of a different recall task, Martin et al. also increased the number of items within each list in Experiment 2: they used 20-item lists in the second experiment instead of the ten-item lists used in the first experiment because they were concerned with the possibility of ceiling effects occurring if only ten-item lists were used in the free recall task.

Results from the second experiment showed no effect in early list positions for recall of lists produced by multiple talkers. However, for terminal list positions, recall was actually observed to be better for lists produced by multiple talkers than for lists produced by single talkers. Martin et al. drew several conclusions from the pattern of results obtained in Experiment 1 and Experiment 2. First, they concluded that requiring subjects to encode order information in the serial recall experiment was probably the most important factor responsible for the differences in recall between lists produced by single and multiple talkers for early list items. Second, they concluded that the enhanced recall of late list items in multiple talker lists in the free recall task may have been due to the distinctiveness or discriminability of the individual list items in auditory STM. The different voices for each item served as an additional cue that increased the discriminability of words in the latter part of each list. These additional cues would help keep these items distinct from each other during maintainance rehearsal prior to recall. Martin et al. reasoned that requiring subjects to encode order information in the serial recall task may have impaired the use of acoustic cues unique to specific talkers, thus resulting in no facilitory effect for items from the latter part of the multiple talker lists. This seems very likely, because, in the multiple-talker lists, the voices changed from trial-to-trial and from list-to-list so subjects could not use the voice information as a retrieval cue for the item.

Martin et al. conducted two other experiments using the same stimuli as used in Experiments 1 and 2 but utilizing slightly different experimental procedures. In Experiment 3, a memory preload task was used to increase capacity demands on STM, therefore causing performance in the primary task, a serial recall task, to decline (Baddeley & Hitch, 1974). The preload task consisted of having subjects retain in active working memory three or six digits presented visually prior to the presentation of the word lists. Subjects were required to recall the visual digits in the order in which they were presented, followed by recall of the spoken words. Martin et al. predicted that the addition of the preload task should cause a greater decrement in recall of early list items from word lists produced by multiple talkers than for word lists produced by a single talker. Results indicated that as the memory preload was increased, the difference between the two types of lists in early list positions did not increase. However, recall of the digits used in the preload task did show an effect of increased preload, plus an overall effect of single versus multiple talkers. These results showed that the ability of subjects to recall the digits was adversely affected by the subsequent presentation of word lists produced by multiple talkers as compared to the same lists produced by a single talker. Thus, there was strong evidence for an effect of speaker variability on recall.

Finally, in a fourth experiment, Martin et al. investigated the possibility that the effects observed in the earlier experiments were due to the impairment of retrieval processes from LTM rather than encoding and/or rehearsal processes. In the case where a single talker produces a list of words, the memory trace of the final words from the list that remain in auditory STM preserve the unique characteristics of the talker's voice. Therefore, the cues present in STM associated with the talkers voice could facilitate the recall of earlier list items from LTM because the same cues were associated with that talker's voice during encoding along with item information. In contrast, it is unlikely that when recalling lists produced by multiple talkers that subjects would be able to take advantage of such acoustic cues for list-final items to aid the retrieval of initial list items. To test this hypothesis, Martin et al. used a retroactive interference task to attenuate the effect of STM on recall by requiring subjects to engage in an arithmetic task between hearing the list of words and recall of the list items (Peterson & Peterson, 1959). The duration of the arithmetic task ranged from four to twelve seconds. Results indicated that the arithmetic task did not affect differences in recall at initial list positions for either single or multiple talker lists. The only effect of the arithmetic task was to reduce recall for late list items as the duration of the interpolated activity increased. This finding is consistent with earlier findings in the literature (Peterson & Peterson, 1959). Furthermore, the interference task should have prevented the use of any acoustic cues that would have otherwise been present in STM to facilitate recall in the single talker condition. Instead, recall of early list items in the single talker condition appeared to be unimpaired, even at the longest interference interval. Therefore, Martin et al. concluded that recall of early list items was independent of STM processes that might provide talker cues for the retrieval of list items presented earlier. Rather, the original hypothesis that a normalization process is required for multiple talkers remained the most plausible account of the pattern of data observed across all four experiments.

The work of Martin et al. has illuminated a number of aspects of the differences in recall observed between lists of words produced by single and multiple talkers. However, the findings have also raised several questions. The present investigation was designed to further examine some of the issues raised by this earlier work. In particular, two aspects of Martin et al.'s study were the focus of the present investigation. First, the differences that Martin et al. obtained between single and multiple talker lists using a serial recall task and a free recall task bear further consideration. It is unclear what caused the different patterns of recall performance between the two tasks because in addition to using two different recall tasks, list length was not controlled. Although Martin et al. interpreted the differences between the two recall tasks as a function of whether or not subjects were required to encode order information, explanations related to differences in list length cannot be eliminated. Therefore, in Experiment 1, we examined recall of lists of words produced by single and multiple talkers using a free recall task and a serial recall task in which list length was held constant at ten items. Another problem with the Martin et al. study involved the repetitions of the same word lists four times during the course of their first experiment. In contrast, in their second experiment, subjects heard each word list only once, raising the possibility that this manipulation may have also affected the differences they observed in comparing results from the free and serial recall tasks. Thus, a second goal of Experiment 1 was to also control for this possible confounding.

In Experiment 2, we examined a manipulation designed to improve recall of early list items in lists produced by multiple talkers. If talker variability was the major factor responsible for impaired recall in early list positions, then reducing item variability seemed like a reasonable manipulation that might improve recall. One way to reduce the item-to-item variability in the talker's voice would be to inform subjects that they will be presented with lists of words produced by different talkers but the talkers would always be in the same order on the list. If the order of a particular talker's voice was constant across list presentations, then the detrimental effect of talker variability might be reduced because the voice was mapped consistently to the same serial position on each list. In Experiment 2, we presented subjects with lists of words produced by multiple talkers in which the order of the voices remained constant across each list. Talker 1 always produced items in list position 1 while talker 2 always produced items in list position 2, etc. Instead of varying randomly from item-to-item and list-to-list, a given talker always produced items in the same position in the list.

Experiment 1

In Experiment 1, we compared performance using free and serial recall tasks for word lists produced by single and multiple talkers. Martin et al. (1987) found that patterns of recall between the two conditions varied depending on the recall task used, and concluded that memory for the order information required in serial recall was the primary factor responsible for the differences they observed between the two tasks. Because list length also differed between the two tasks, a possible confounding between type of task and list length existed. Therefore, the major goal of Experiment 1 was to see if the results obtained by Martin et al. could be replicated if list length was held constant in the two tasks.

A secondary goal of Experiment 1 was to examine the effects on recall of several stimulus properties related to the acoustic-phonetic confusability of each word. The motivation for considering such variables was the recent work of Luce (1986) who has investigated the effects of neighborhood similarity on the recognition of spoken words. Neighborhood similarity is a measure of how similar one word is to other words in the mental lexicon based on common sound patterns. Some words, such as "dot", for example, come from high density neighborhoods of the lexicon where there are many words that have sound patterns that are similar to "dot" in the mental lexicon. Other words, such as "deluge", for example, come from much lower density neighborhoods where there are only a small number of words which have sound patterns similar to "deluge" in the mental lexicon. Luce (1986) found that, taken together with word frequency, neighborhood density could predict subject's performance in various word recognition tasks. Subjects performed best when presented high frequency words from low density neighborhoods whereas their performance was worst for low frequency words from high density neighborhoods. Luce explained these results in terms of the competition that a given word in the lexicon has from other words that sound similar to that word, how frequently the word is encountered, and also the frequency of similar-sounding words.

In the present experiment, the words in the lists were chosen on the basis of Luce's (1986) findings. Half of the lists contained "easy" words -- high frequency words from low density neighborhoods, while the other half of the lists contained "hard" words -- low frequency words from high density neighborhoods. Thus, two sets of stimuli were generated, one which contained relatively confusable items and one which contained less confusable items. The degree of within-list confusability provided an additional way to determine the extent to which acoustic-phonetic discriminability could help maintain the distinctiveness of items remaining in STM at the time of recall. Specifically, we were interested in determining whether recall of late list items would be enhanced for "easy" words because they were less confusable compared to "hard" words. We were also interested in whether confusability would also affect early list positions due to differences at the time of encoding. Would "easy" words be recalled better than "hard" words? Finally, we wanted to determine whether the confusability manipulation would interact with talker variability. Experiment 1 was designed to answer all these questions.

Method

Subjects. Eighty-eight students enrolled in an introductory psychology course at Indiana University in Bloomington served as subjects. Subjects received course credit for their participation. All were native speakers of English and all reported no history of a speech or hearing disorder at the time of testing.

Stimuli. The stimuli were obtained from the same source used by Martin, Mullennix, Pisoni, and Summers (1987), a large digital database of spoken materials recorded by several different talkers. The original source of the monosyllabic words was the Modified Rhyme Test (House, Williams, Hecker, & Kryter, 1965). In the present experiment, only a subset of the original 300 words were used. The words used in the present experiment were chosen according to several structural criteria based on computational analyses of the database. First, the words were ranked according to their frequency of occurence using the frequency norms from Kucera and Francis (1967). Second, the words were also ranked according to their phonetic confusability as determined by a one-phoneme substitution metric (Luce, 1986). Words that came from high-density neighborhoods in the lexicon had many similar-sounding confusable words whereas words that came from low-density neighborhoods had fewer similar-sounding words. Third, words were also ranked according to neighborhood frequency, a measure of the average frequency of the words that are in a lexical neighborhood. Using these three criteria, two sets of words were chosen for use in the present experiment. One set, the "easy" words, consisted of high frequency words selected from low density neighborhoods with low frequency neighbors. The other group of words, the "hard" words, were low frequency words from high density neighborhoods with high frequency neighbors. The two different word sets paralleled results from experiments by Luce (1986) who showed that "easy" words are identified more quickly and accurately than "hard" words. One final criterion used to select the words was subjective familiarity; all of the words chosen for use in the present experiment were rated as highly familiar to subjects based on norms collected in an earlier study (Nusbaum, Pisoni, & Davis, 1984). After applying these four criteria, the "easy" and "hard" word sets each contained 50 items. These words were then used to generate 10 lists, five lists containing "easy" words and five lists containing "hard" words. Each list contained ten words.

After creating the lists of words, digitized files containing tokens of each word were selected from the database. One set of tokens was chosen from utterances produced by a single male talker; these stimuli were used in the single talker condition. Another set of tokens was chosen so that for each of the ten words contained in a list, each word was chosen from utterances produced by a different talker; these stimuli were used in the multiple talker condition. In the multiple talker condition, the same ten talkers, five males and five females, were used in all ten lists of words. Thus, one set of stimuli consisted of words produced by one talker while the other set of stimuli consisted of the same words produced by ten different talkers. The present experiment used the same set of talkers as used in Martin et al. (1987). All of the speech stimuli were originally recorded on audio tape and then digitized with a 12-bit analog-to-digital converter using a PDP 11/34 computer. RMS amplitude of all stimulus tokens was equated using a signal processing package.

Procedure. Subjects were tested in groups of two to six in a quiet testing room used for speech perception experiments. Each subject was seated at an individual booth with a desk. Stimuli were presented over matched and calibrated TDH-39 headphones at 75 dB SPL as measured by a VTVM. A PDP 11/34 computer was used to present the stimuli and to control the experimental procedure in real-time. The digitized stimuli were reproduced using a 12 bit digital-to-analog converter and were then low-pass filtered at 4.8 kHz.

All subjects were tested under the same conditions with the exception of the type of recall task used. Subjects first heard a 500 ms 1000 Hz warning tone indicating that a list of words was about to be presented. Then, a list of ten words was presented with an inter-word interval of 1500 ms. A tone was presented after the list had ended to indicate the beginning of the recall period. Subjects had 90 s to perform the recall task. The end of the recall period was indicated by the presentation of a third tone. Subjects were instructed to recall as many items as they could during the recall period and to use the entire period for recall. Subjects in the free recall condition were told to recall items with no restrictions on the order of recall. Subjects in the serial recall condition were told to recall items in the same order as they were presented in the lists. Subjects in both conditions wrote their responses in specially prepared answer booklets using pen or pencil.

Recall task and talker condition were between-subject variables in the present experiment. Forty-four subjects were tested using the free recall task and 44 subjects were tested using the serial recall task. For each task, half of the subjects listened to lists produced by a single talker while the other half of subjects listened to lists produced by multiple talkers. The same word lists were heard by all subjects; only the talkers and recall condition varied between subjects. The order of presentation of items within a list varied randomly from session to session. The lists themselves were presented in the same order in all conditions of the experiment; the presentation of lists alternated between those containing "easy" and those containing "hard" words.

Results

The data were scored according to the following criteria. In the free recall condition, responses were scored as correct if they were either the target word or some phonetically equivalent spelling regardless of their position in the list. In the serial recall condition, responses were scored as correct if, and only if, they were in the same serial order as the item presented on the list.

Figure 1 shows the percentage of correctly recalled words as a function of serial position for both the free recall condition (top panel) and the serial recall condition (bottom panel). Each graph shows data for both single and multiple talker conditions averaged over the two types of word lists.

Insert Figure 1 about here

As expected, in both free and serial recall conditions, an effect of serial position was present as shown by increases in the percentage of correctly recalled items at the beginning and end of the list. Also, each panel shows the effects of the talker manipulation. In the free recall condition, shown at the top, the effect of talker appears to be present only in the early positions of the list. These data also appear less systematic than the serial recall data. Inspection of the serial recall condition shows two interesting effects. First, as anticipated from previous work, recall of items produced by multiple talkers is worse than recall of items produced by a single talker in the first half of the list. This replicates the earlier findings reported by Martin et al. However, there is also an effect of talker variability on recall of items in the second half of the list. Now recall of items produced by a single talker is actually worse than recall of items produced by multiple talkers in the second half of the list. This reversal was unexpected and may reveal important differences in the nature of the rehearsal process for these items at different points in the serial position curve.

An analysis of variance was used to assess the effects of recall task (free versus serial), talker (single versus multiple), position in list (1-10), and type of word ("easy" versus "hard"). A significant main effect for recall task was obtained. Overall, recall performance was better in the free recall task than in the serial recall task $F(1, 84)=28.55$, $p<0.001$. A significant main effect for list position was also obtained, $F(9,756)=87.1, p<0.001$, reflecting the overall serial position effect across recall tasks, talkers, and types of word lists. A significant main effect was also obtained for word type, $F(1,84)=221.84$, $p<0.001$, indicating that overall, "easy" words were recalled better than "hard" words.

Several significant two-way interactions were also obtained: First, a significant interaction between word type and recall task was obtained, $F(1,84)=5.77$, $p<0.05$. The difference in recall between "easy" and "hard" words was greater in the free recall condition than in the serial recall condition. More specifically, the percentage of "easy and "hard" words

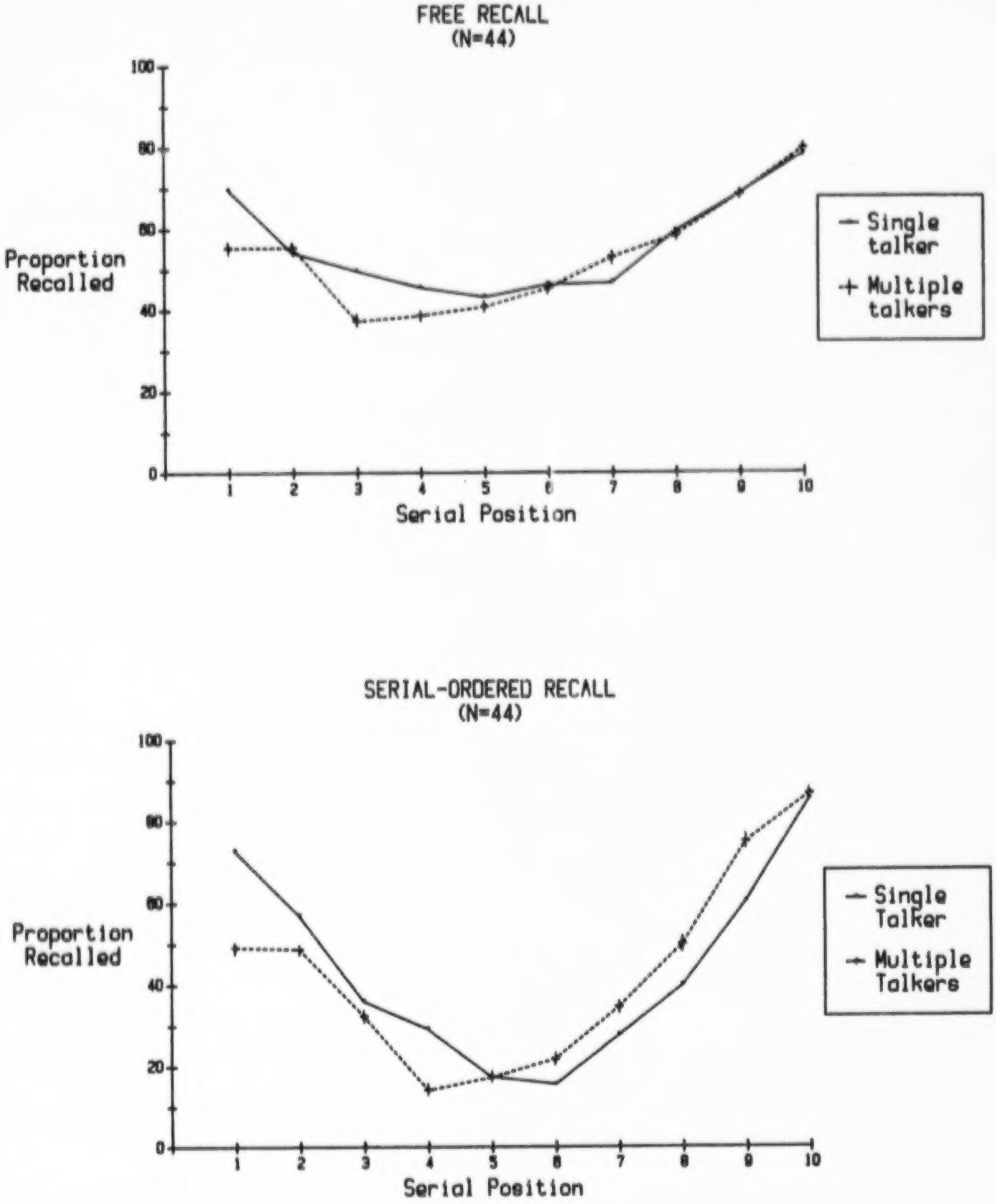


Figure 1. Percentage of correctly recalled items for single and multiple talkers as a function of serial position in free (top panel) and serial (bottom panel) recall tasks.

recalled in the free recall task was 62.5% versus 46.9%, respectively. In contrast, the percentage of "easy and "hard" words recalled in the serial recall task was 49.2% versus 37.9%, respectively. Second, there was a significant interaction between serial position and recall task, F(9, 756)=9.5, p<0.001. This was due primarily to the lower recall in the middle of the list in the serial recall condition compared to recall in the same region in the free recall condition. Third, there was a significant interaction between serial position and talker, F(9, 756)=5.09, p<0.001. This result is due to the reversal of the effects shown in Figure 1 in which recall of items produced by multiple talkers in the first half of the list is worse than for recall of items produced by single talkers. This effect is reversed in the second half of the list. Post-hoc tests indicated that the only position in the list where recall was significantly different for single and multiple talker conditions was in the first position. Despite the lack of a statistically significant difference in recall performance between single and multiple talker conditions at other list positions, the overall pattern of results is similar to that obtained by Martin et al. (1987). Since the overall pattern of results with respect to talker variability found in the present experiment is similar to that found by Martin et al., the lack of a statistically significant effect at other list positions can probably be attributed to an insufficient number of observations per cell.

Fourth, there was a significant interaction between word type and serial position, F(9, 756)=4.6, p<0.001. Figure 2 shows the mean recall for "easy" and "hard" words as a function of list position summed over recall task. The interaction between these two variables is due to the reduced recall of "hard" words compared to recall of "easy" words in the early part of the lists. In other words, although recall of "hard" words was always consistently worse than recall of "easy" words, the largest decrement in recall for the "hard" words was in the early list positions.

Insert Figure 2 about here

The ANOVA revealed one further effect, a significant three-way interaction among word type, recall task, and talker, F(1,84)=7.88, p<0.01. Figure 3 shows the percentage of "easy" and "hard" words recalled in the single and multiple talker conditions for the free and serial recall tasks. The top panel of Figure 3 shows data from the free recall task whereas the bottom panel of Figure 3 shows data from the serial recall task. In the free recall task, an interaction occurred between word type and talker; for "easy" words, there was no difference in recall for lists produced by multiple talkers compared to lists produced by the single talker whereas for "hard" words, recall was better for single talker lists than for multiple talker lists. In the serial recall task, recall appears to be similar for single talker lists and multiple talker lists, regardless of word type.

Insert Figure 3 about here

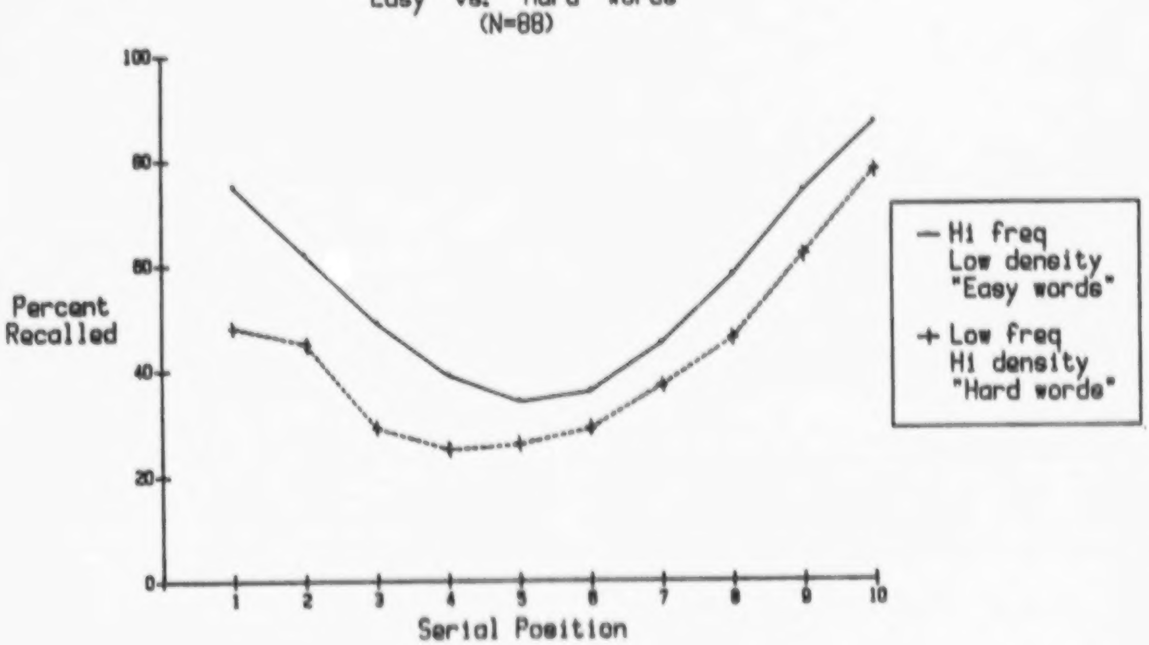
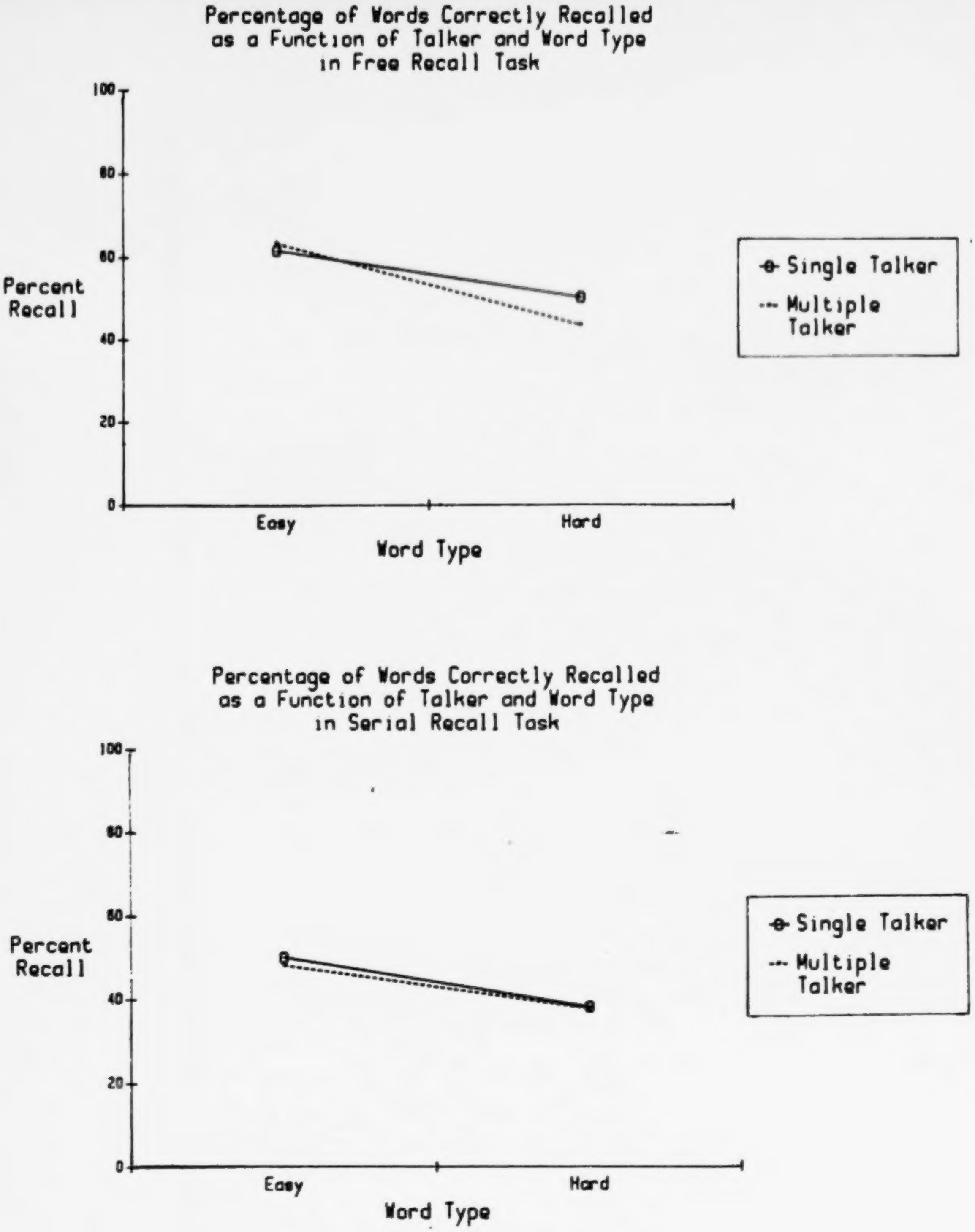


Figure 2. Percentage of correctly recalled words for "easy" and "hard" words as a function of serial position.



Percentage of Words Correctly Recalled as a Function of Talker and Word Type in Free Recall Task

Percentage of Words Correctly Recalled as a Function of Talker and Word Type in Serial Recall Task

Figure 3. Percentage of words recalled from "easy" and "hard" lists produced by single and multiple talkers in the free recall task (top panel) and the serial recall task (bottom panel).

Discussion

Overall, the pattern of results revealed several interesting effects. First, a more consistent effect of talker variability on recall was obtained using the serial recall task than using the free recall task. However, the largest effect for talker voice was found in the free recall task in the first position. This result differs from that obtained by Martin et al. (1987) who found that for early list positions, the serial recall task showed differences between single and multiple talker lists whereas the free recall task showed no difference in recall for the two types of lists. These results suggest that the difference between the length of the word lists used in Martin et al.'s experiments may have accounted for some of the differences they observed between the free and serial recall tasks. When list length was held constant, as in the present experiment, talker variability appears to affect recall in both tasks.

However, another difference in methodology may also account for some of the differences observed in the present experiment. In Martin et al.'s Experiment 1, list items were repeated four times. That is, each ten-word list was presented four times in a different order during the course of the experiment. In the present experiment, the same list items were presented only once. The repetition of the same stimulus items during the experiment undoubtedly had some effect on memory processes, resulting in an improvement in encoding due to the repeated exposure to the same items. It is unclear whether this effect would be greater for single or multiple talker lists.

The results of the present experiment and the previous results of Martin et al. display similar patterns in overall recall performance for both free and serial recall. The present experiment shows that two different effects may be present in recall of lists of words produced by single and multiple talkers. In early list positions, recall is consistently better for single talker lists than for multiple talker lists. In contrast, for late list positions, recall is better for multiple talker lists than for single talker lists. These effects were found more consistently when a serial recall task was used than when a free recall task was used. Martin et al. suggested that these two effects reflect the operation of two fundamentally different memory mechanisms. The decrease in recall observed in early list positions for items from multiple talker lists was ascribed to initial encoding difficulties due to the demands of talker normalization which in turn interfered with the rehearsal and subsequent transfer of information to LTM. The enhanced recall in late list positions for items from multiple talker lists was ascribed to the facilitory effect of different voices in maintaining the distinctiveness of individual items during active rehearsal in auditory STM.

The confusability variable appeared to have its greatest effect at early list positions, implicating encoding processes responsible for transfer of the items from STM to LTM. Similar results were obtained by Sumby (1963) and Raymond (1969): In a free recall task, recall was lower in early list positions for low frequency words than for high frequency words. The pattern of results found in both of these earlier experiments was used primarily as evidence to support a two-store model of memory. No attempt was made by either Sumby or Raymond to explain their results in terms of some specific mechanism. In retrospect, the use of frequency as a variable by Sumby and Raymond was confounded with other stimulus variables such as neighborhood density, neighborhood frequency, and familiarity, therefore making it difficult to identify the factors that were actually responsible for the effects they obtained.

Experiment 2

The results of Experiment 1 replicated the essential findings of Martin et al. (1987). That is, recall of early list items was impaired for lists produced by multiple talkers compared to the same list produced by single talkers. Furthermore, facilitory effects of words produced by multiple talkers compared to lists produced by single talkers was found at late list positions. Taken together, these two findings suggest that subjects may be attempting to use talker-specific cues to help encode item and order information in the serial recall task. One way to investigate the nature of these effects is to see if the results would be affected by different types of list manipulations. The purpose of the present experiment was to determine if we could improve recall performance by maintaining a constant and predictable ordering of the talkers in the multiple talker condition. We predicted that by presenting listeners with words from multiple talkers in a consistent order that was maintained from list to list, subjects would be able to use the correlated and redundant speaker-specific cues to help encode item and order information. We expected to find the largest effects of this manipulation in early serial positions. It was unclear how the manipulation would affect other list positions, although we assumed that performance in the other list positions would be at least as good as that obtained in Experiment 1 for the randomly-ordered multiple talker condition. A consistent mapping of talker to list position could only produce increases in recall. The question we were interested in was whether this improvement would be selective at early list positions.

Method

Subjects. A total of 22 subjects were obtained from the same source as in Experiment 1. All subjects reported no history of a speech or hearing disorder.

Stimuli. The stimuli were the same stimuli as used in Experiment 1.

Procedure. Subjects in Experiment 2 were tested in a serial recall task and were presented only with multiple talker lists. The major change in the procedure was that, whereas the order of talkers in the multiple talker condition of Experiment 1 varied randomly from list to list, the order of talkers in Experiment 2 was held constant from list to list. That is, for all lists, the talkers always appeared in the same serial order although they produced different items. However, each group of subjects tested was presented with the talkers in a different random order. At the completion of each experimental session, subjects were also asked to recall the gender of the talkers. This was done to see if subjects attended to the consistent ordering of the talkers across the lists and also to see if there was any correlation between recall of the order of the talkers' voices.

Results

The data were scored according to the same criteria used in the previous serial recall task. Figure 4 shows the percentage of correctly recalled words as a function of serial position. For purposes of comparison, data from the single and multiple talker conditions of Experiment 1 are also shown in this figure.

Insert Figure 4 about here

Inspection of Figure 4 indicates that the recall of items in the consistent order condition is generally better than recall in the random order condition although it is not quite as good as recall in the single talker condition. This manipulation was successful in increasing the recall of items produced by multiple talkers compared to when the same voices are just randomly ordered from list to list. However, the manipulation did not produce recall performance equal to the performance obtained with the single talker.

An ANOVA comparing recall of the random order and the consistent order conditions was carried out. Significant main effects were found for confusability, $F(1, 42)=84.01$, $p<0.0001$, and serial position, $F(9, 378)=72.95$, $p<0.0001$. However, there was no significant main effect of the talker manipulation. A significant two-way interaction between position and talker was obtained, $F(9, 378)=2.69$, $p<0.005$. This interaction was due to the improvement in recall for the consistent order condition compared to the random order condition, especially at early and middle list positions. In late list positions, recall of the consistent order lists does not appear to be different from the random order lists. In short, the facilitory effects of maintaining a consistent order for lists produced by multiple talkers appear to be selective in nature and limited to early and middle list positions.

As described above, in addition to recall of the word lists, subjects were also asked to recall the gender of the talkers producing the word lists after completion of the main part of the experiment. We reasoned that there might be a relation between a subject's overall performance on the recall task and their ability to successfully use speaker-specific cues to improve recall performance. Any such relationship between recall performance for words and specific memory for a talker's voice would, in all likelihood, be tacit, a relationship that subjects learned incidentally while carrying out the primary task. In order to assess the possibility that those subjects who were successful in recalling the order of the talkers at the conclusion of the experiment also demonstrated better recall of the word lists, a further analysis was performed. A Pearson product-moment correlation showed no relationship between a subject's recall of the words and recall of the order of the talkers producing the words ($r = 0.045$). Two further correlations were calculated in order to assess the relationship between "easy" and "hard" words and talker recall. However, for both types of words, no significant correlation was obtained between recall of the word lists and recall of talker order. Thus, no obvious relationship between recall and overt memory for talker characteristics was observed in the present experiment.

Discussion

The results of Experiment 2 show that by consistently ordering lists of words with respect to the talker's voice, recall performance was improved compared to when the same lists were produced by talkers that changed from trial to trial. Improvement was localized in early and middle list positions, while little effect was found in late list positions.

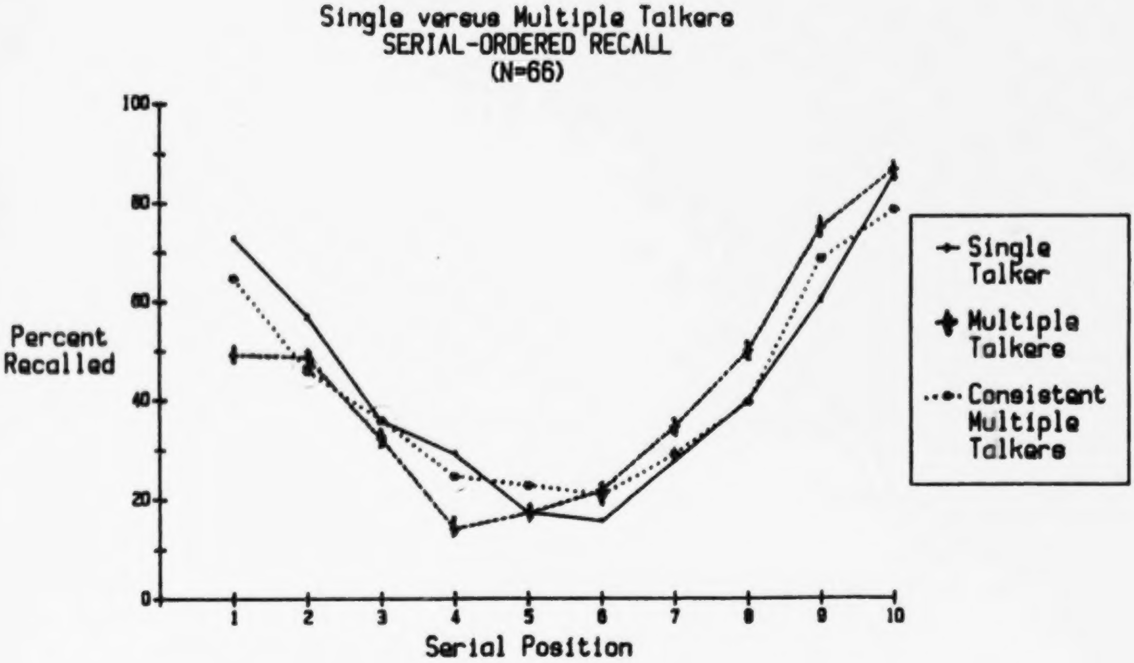


Figure 4. Percentage of correctly recalled words as a function of serial position from the consistently-ordered multiple talker condition. For purposes of comparison, data from the serial recall single and multiple talker conditions of Experiment 1 are also shown.

A theoretical account of the results obtained when talker order is held constant for lists produced by multiple talkers would appear to involve the use of talker-specific cues that were encoded along with each item. The repeated presentation of lists in which the talker was consistently mapped to a list position enabled subjects to use that information at retrieval as a cue to recall items in the order in which they were presented. Compared to the condition in which talker order was consistent, the random order of presentation does not provide such retrieval cues to be used during recall since talker order varied from list to list. Subjects could not use this information to encode the order and items because it was not correlated with the to-be-remembered information.

General Discussion

The results of Experiment 1 showed that for early list positions, recall of words produced by single talkers was better than recall of words produced by multiple talkers. This finding replicated earlier work in our laboratory by Martin et al. (1987). In contrast, for late list positions, recall of words produced by multiple talkers was actually better than recall of words produced by single talkers. The effect was more consistent in a serial recall task than in a free recall task. Thus, the results of Experiment 1 were in general agreement with the essential findings of Martin et al. (1987). Manipulation of confusability produced effects on recall that varied with serial position and talker. The results of Experiment 2 showed that making talker order consistent in the multiple talker condition caused a small but selective improvement in recall in early and middle list positions. This manipulation enabled talker-specific cues to be encoded together with item information and therefore facilitated recall from LTM when talker order was used as a recall cue.

Taken together, the results of Experiments 1 and 2 provide additional information about the recall of spoken word lists. First, lists produced by multiple talkers require additional processing capacity for encoding beyond the capacity required for lists produced by single talkers. Presumably, the extra processing capacity required by multiple talker lists is a consequence of speaker normalization processes, a mapping of the characteristics associated with individual talkers on to a more abstract representation used to make contact with the listener's lexicon (see Mullennix et al., 1987). Allocation of processing capacity for speaker normalization results in less capacity available for rehearsal and other encoding processes. Therefore, recall of items from early and middle positions is reduced for multiple talker lists compared to single talker lists. In contrast, for late serial positions, lists produced by multiple talkers may be recalled more accurately than lists produced by a single talker. In this case, the distinctiveness of the items in the multiple talker condition may aid in keeping the items separate and more distinctive in auditory STM and therefore increasing the likelihood of correct recall. The effect of maintaining a consistent talker order in the multiple talker condition arises from the operation of the same memory mechanisms. By maintaining talker order in the multiple talker condition, encoding of early list items may be affected less by the variability due to different talkers. Furthermore, consistently ordered lists also aid retrieval through the use of talker-specific information as a cue for accessing talker information that was associated with the item during encoding. The effect of the consistent talker manipulation tends to be limited to initial and middle list positions since recall from these list positions is primarily a function of retrieval from LTM and therefore subject to use of the talker-specific retrieval cues. In contrast, the rehearsal processes that maintain items in STM are not likely to be affected any more in lists that are consistently ordered compared to lists that are randomly ordered.

The results of the present investigation also indicated that the confusability of stimulus items influenced overall recall, with the exact nature of the effect depending on list position and talker variability. In general, confusability affected early list positions more than later list positions, again implicating encoding processes. The role of confusability in the recall of spoken word lists needs to be examined more in future research, especially with regard to how this manipulation interacts with the acoustic information associated with talker identity. In future experiments dealing with confusability and recall of spoken word lists, the lists should contain words with more extreme values of frequency, neighborhood density, and neighborhood frequency in order to maximize the effect of these variables on recall performance.

Finally, the results of the present investigation can be viewed as further evidence for traditional two-store models of memory in which separate STM and LTM systems are posited (eg., Atkinson & Shiffrin, 1968). The interpretation of the present results is entirely consistent with this class of models. However, alternative conceptions of how information is stored and recalled from memory are also possible. For example, Greene (1986a; 1986b) has argued recently that the STM - LTM distinction is inappropriate. Of special relevance to the present set of experiments is Greene's (1986a) investigation of word frequency and its effect on recall. He found that, similar to Sumby (1963), Raymond (1969), and the results of the present experiment, low freqeuency words tended to be recalled more poorly than high frequency words in all list positions except late positions. The presence of this effect had been used earlier as evidence for the existence of separate STM and LTM systems since only early and middle list positions were affected. However, Greene showed that this frequency effect could also be obtained even when a continuous distractor task was used during list presentation. According to the traditional STM - LTM view, the distractor task should have occupied the limited capacity STM system, eliminating any recency effect. Yet, the effect of the frequency manipulation remained, suggesting that the traditional explanation of recency as a STM phenomenon was incorrect. Thus, it is likely that an account of the results of the present set of experiments does not have to rely exclusively on the STM - LTM distinction even though that is the way in which we have chose to present them here.

In summary, the present set of experiments demonstrated that recall of lists of spoken words produced by single and multiple talkers differed depending upon list position. In early list positions, recall was better for single talker lists whereas in late list positions, recall was better for multiple talker lists. Furthermore, recall of single and multiple talker lists was also shown to be dependent on the acoustic-phonetic confusability of the words within the lists. Finally, the results also indicated that recall in early list positions of words from multiple talker lists could be improved by consistently ordering the talkers across lists thus providing the subject with additional speaker-specific redundant cues which can be used for encoding items and order information. Overall, the results of the present experiments show that talker variability affects initial encoding and/or rehearsal processes and the transfer of spoken items to long-term memory. Normalization processes required to compensate for talker variability appear to require additional processing resources and demands which apparently also affect other processes related to both short- and long-term memory. To our knowledge, this is the first time both increases and decreases in recall performance were observed with the same talker manipulation at different serial positions.

These findings also provide further support for traditional two-process models of memory involving both short- and long-term components that have fundamentally quite different dynamics and operating principles.

References

Atkinson, R.C., & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence & J.T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89-105). New York: Academic Press.

Baddeley, A.D., & Hitch, G.J. (1974). Working memory. In G.H. Bower (Ed.) *The psychology of learning and memory* (Vol. 8). New York: Academic Press.

Craik, F.I.M., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*, *26*, 274-284.

Greene, R. L. (1986a). A common basis for recency effects in immediate and delayed recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 413-418.

Greene, R. L. (1986b). Sources of recency effects in free recall. *Psychological Bulletin*, *99*, 221-228.

Kucera, F., & Francis, W. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

Luce, P.A. (1986). Neighborhoods of words in the mental lexicon. *Research on Speech Perception, Technical Report No. 6*. Bloomington, IN: Indiana University.

Luce, P.A., Feustel, T.C., & Pisoni, D.B. (1983). Capacity demands in short-term memory for synthetic and natural word lists. *Human Factors*, *25*, 17-32.

Martin, C.S., Mullennix, J.W., Pisoni, D.B., & Summers, W.V. (1988). Effects of talker variability on recall of spoken word lists. *Research on Speech Perception Progress Report No. 13*. Bloomington, IN: Indiana University.

Mattingly, I.G., Studdert-Kennedy, M., & Magan, H. (1983). Phonological short-term memory preserves phonetic detail. *Journal of the Acoustic Society of America*, *73*, S6.

Mullennix, J.W., Pisoni, D.B., & Martin, C.S. (1987). Some effects of talker variability on spoken word recognition. *Research on Speech Perception Progress Report No. 13*. Bloomington, IN: Indiana University.

Peterson, L.J., & Peterson, M.J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*, 193-198.

Raymond, B. (1969). Short-term storage and long-term storage in free recall. *Journal of Verbal Learning and Verbal Behavior*, *8*, 567-574.

Shiffrin, R.M. (1976). Capacity limitations in information processing, attention, and memory. In W.K. Estes (Ed.) *Handbook of learning and cognitive processes* (Vol. 4). Hillsdale, NJ: Erlbaum.

Sumby, W. H. (1963). Word frequency and serial position effects. *Journal of Verbal Learning and Verbal Behavior*, *1*, 443-450.

III. INSTRUMENTATION AND SOFTWARE DEVELOPMENT

[RESEARCH ON SPEECH PERCEPTION Progress Report No. 13 (1987) Indiana University]

SAP: A Speech Acquisition Program for the SRL-VAX*

Michael J. Dedina

Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, IN 47405

---

* The development of this software was supported, in part, by NIH Research Grant NS-12179-11 and, in part, by the Armstrong Aerospace Medical Research Laboratory Contract No. AF-F-33615-86-C-0549 to Indiana University in Bloomington.

SAP: A Speech Acquisition Program for the SRL-VAX

Introduction

SAP (Speech Aquisition Program) is a program used to construct databases of digitized utterances. Traditionally, we have collected utterances by first tape recording lists of spoken words or sentences using traditional analog techniques and then later digitizing and editing the utterances into separate files using a digital waveform editor. This method has usually been adequate for our needs, but it can be very time-consuming when a very large number of utterances must be collected and analysed. SAP digitizes speech directly as it is being spoken, creating a separate digital file for each utterance. Thus, SAP not only saves time previously spent digitizing and editing speech, but it also provides the advantage of eliminating the degradation to the acoustic signal resulting from intermediate storage on analog audio tape and playback.

This paper describes SAP and also serves as a user manual. Program specifications and hardware requirements are briefly discussed, then the operation of the program is described. For the benefit of users of SAP, there is a section describing the files used by the program, and finally, a section describing the dialog between the experimenter and the program.

Specifications and Hardware Requirements

SAP was written in FORTRAN by Moshe Yuchtman and Mike Dedina and currently runs on the SRL VAX-11/750 under VMS. The major hardware component supporting analog-to-digital conversion is a dual-channel DSC-200 Audio Data Conversion System from Digital Sound Corporation. The DSC-200 provides 16-bit A-D and D-A resolution, with selectable 4.8 KHz and 9.6 KHz filters for sampling rates of 10 KHz or 20 KHz. SAP is run and controlled from an ILS graphics workstation consisting of a Retrographics graphics display terminal and a DSC-240 Audio Control Console, which provides the interface for connecting microphones and speakers to the DSC-200. SAP controls the DSC-200 via subroutine calls to the DSC Audio Subsystem Portable Interface Library (ASPLIB).

Program Operation

From the Talker's Viewpoint

The talker is seated at the "talker station", an isolated sound-attenuated acoustic chamber equipped with a VT100 terminal connected to the VAX, and a microphone mounted on a set of headphones. The headphones are connected to an analog audio rack equipped with a white noise generator, allowing the experimenter to present noise to the talker during the session, if desired. The microphone is connected to an input of the DSC-240 audio console. The headphone/microphone configuration helps to keep a constant distance of about three inches between the microphone and the talker's mouth. SAP prompts the talker with character strings presented on the VT100. For each trial in the session, SAP presents the word or sentence to be spoken on the talker's terminal in large characters, then immediately initiates a sampling interval during which the DSC-200 digitally samples the auditory input from the microphone. The experimenter specifies the length of the sampling intervals at the beginning of the session, but the talker is able to terminate the interval and move on to the next trial immediately after speaking a word or sentence by hitting the carriage return. This makes the session self-paced and serves to speed up collecting data in an on-line mode.

From the Experimenter's Viewpoint

The experimenter runs SAP from an audio workstation which is located in the same room as the talker station. Recording levels are set via LED indicators on the DSC 240. SAP first queries the experimenter for information concerning file names, the number of trials, repetitions, and blocks, and the length of the sampling intervals. An example of this dialog between the experimenter and SAP is included at the end of this report. SAP then prompts the talker to hit the carriage return on his terminal when he is ready to begin the session.

The experimenter is able to monitor the session and take corrective action if a word is mispronounced by the talker. SAP displays information about the current trial on the experimenter's screen including the trial number and the word or sentence being spoken. In addition, the DSC-240 permits concurrent monitoring of the talker's utterances over a speaker at the experimenter's workstation. If the experimenter notices that the talker did not speak a particular word clearly, he can back up any number of trials by hitting the appropriate key on his terminal.

Files

Text File (Stimulus Materials)

The experimenter must provide SAP with a text file, which contains the words or sentences to be presented to the talker and specifies part of the file name for each utterance. The text file is an ascii file created with a text editor, and should have the extension TXT. The file contains two lines for each utterance to be spoken by the talker. The first line consists of two characters, which are incorporated into the utterance file name for that utterance (see the next paragraph for an explanation of utterance files). These two characters should uniquely identify the word or sentence. For instance, they might be the first two letters of the word, or a stimulus number. The second line is simply the word or sentence itself. Figure 1 shows a typical text file.

Insert Figure 1 about here

Figure 1
Example Text (Stimulus) File

```
be
bead
de
deed
ge
geed
pe
peed
te
teed
ke
keyed
ba
bad
da
dad
ga
gad
pa
pad
ta
tad
ka
cad
bo
bod
do
dod
go
God
po
pod
```

Figure 1. An example text file which contains the stimulus material to be presented to the talker. The file consists of pairs, one pair for each stimulus word (or sentence). The first item of each pair is a two-letter string which is incorporated into the utterance file name for that word. In this example, it consists of the first two letters of the stimulus word, but it can be any unique two-letter string which identifies the word or sentence. The second item is the actual word (or sentence) which is presented to the talker.

Utterance Files (Speech Signals)

Each digitized utterance is stored by SAP in a disk file as it is being sampled. The size of these files is fixed by the length of the sampling interval, as specified by the experimenter. These utterance files are written in ILS sampled data file format, compatible with the ILS signal processing software package. Thus, the utterances are available immediately after a digitizing session for analysis or processing by the software tools available in the ILS signal processing environment. For example, the experimenter can check an utterance immediately after a session by displaying the utterance as a waveform with ILS's DSP command, or by listening to it with the LSN commmand. Other signal processing operations are readily available as well, using already existing tools.

In accordance with ILS restrictions on file names, sampled utterance files are given names consisting of four alphabetic characters followed by three digits, with no extension. The file names are constructed by SAP in the following manner. The first two letters are the talker's initials, which are provided by the experimenter at the start of the session. The next two letters identify the word or sentence that was spoken; usually they are the first two letters of the word or sentence, but can be anything that uniquely identifies the stimulus. The experimenter specifies these two letters with the text file, as explained above. The first digit in the utterance file name is the condition number that applies to the utterance (e.g., noise vs. quiet), which is provided by the experimenter. Finally, the last two digits of the file name specify the token number, or repetition number, of the utterance. For instance, if an utterance file contained the word DOG, spoken by talker CD in condition 3, and was the twelfth repetition of DOG in that session, the file would be named CDDO312.

Hardcopy File

SAP creates a record of the session by producing a "hardcopy file." The hardcopy file is an ascii file which contains information about the session, which can be printed out and reviewed at a later time. The hardcopy file contains all the parameters entered by the experimenter, along with the date and time the session was run, the stimulus words that were presented, the order in which they were presented, and the names of the files the utterances were stored in.

Sample Dialog

This section illustrates the interaction between the experimenter and SAP by presenting an example session. In this illustration, program dialog is presented in upper case and experimenter responses appear after the arrow prompts (-->). A carriage return is denoted by <CR>. Note that SAP always provides defaults, which are shown in parentheses. A default can be entered by just hitting the carriage return. SAP uses the values from the last time the program was run as the default values. SAP remembers these parameters between sessions by maintaining a file called "SAP.DEF". The experimenter need not be aware of this file. If it doesn't exist, SAP will create it and will use a set of "default defaults" for the current session.

The program signs on and asks for the device name of the talker station terminal, which is TXA6: in our current configuration:

```
SAP V2.2 - SPEECH ACQUISITION PROGRAM
TALKER TERMINAL DEVICE NAME (DEFAULT IS TXA6:)
--> <CR>
```

Here, the experimenter has selected the default by entering a carriage return. Next, the program asks whether the talker display should be in small or large characters:

```
DISPLAY SINGLE OR DOUBLE SIZE (DEFAULT IS DOUBLE)
--> <CR>
```

SAP next asks for the sampling duration and the sampling rate:

```
SAMPLING DURATION IN SECONDS (DEFAULT IS 2)
--> 3
SAMPLING RATE (DEFAULT IS 10000)
--> <CR>
```

Note that in response to the next to last query, the experimenter has overridden the default value. Next SAP asks for the DSC channel number, which determines which of the laboratory's audio consoles will be used in providing input. Currently, our talker workstation uses Channel 2.

```
INPUT CHANNEL (DEFAULT IS 2)
--> <CR>
```

SAP next asks for the text file name, and for information about the talker and experimenter:

```
TEXT (STIMULUS) FILE NAME (DEFAULT IS HVD.TXT)
--> <CR>
EXPERIMENTER NAME (DEFAULT IS ARCHIE)
--> <CR>
TALKER NAME (DEFAULT IS REGGIE)
--> JUGHEAD
TALKER INITIALS (DEFAULT IS RG)
--> JH
```

The program now asks for information about the number of utterances to be collected. Note that the "Number of Stimuli" should be equal to the number of lines in the text divided by two, since there are two lines, a word or sentence and a part of the file name, for each stimulus to be presented.

```
NUMBER OF STIMULI TO BE PRESENTED (DEFAULT IS 10)
--> <CR>
```

Trials are grouped into blocks. SAP asks for the number of blocks and the number of repetitions of each stimulus in each block.

```
NUMBER OF BLOCKS (DEFAULT IS 2)
--> <CR>
```

```
NUMBER OF REPETITIONS (DEFAULT IS 2)
--> <CR>
TOTAL TRIAL COUNT 40.
```

SAP notes the number of trials to be run, 40 in this example. Next SAP asks for the name of the hardcopy file.

```
HARDCOPY FILE NAME (DEFAULT IS TEST.OUT)
--> JHHVD.OUT
```

Before each block, SAP asks for an arbitrary condition label, and a condition number. The program then waits for the talker to hit a carriage return to begin the experiment.

```
ENTER CONDITION LABEL FOR NEXT BLOCK --> 1
WAITING FOR TALKER TO RESPOND...
```

During the course of the session, the experimenter can back up a few trials if he notices that the talker mispronounced a word. This is done by hitting one of the keys "1" through "9" on the experimenter's terminal. The number indicates the number of trials to back up. In addition, the talker can back up one trial by hitting the backspace key on his terminal. As mentioned above, the subject can hit the carriage return immediately after saying a word or sentence in order to terminate the sampling interval and move on to the next trial more quickly.

Conclusion

SAP has been used in the Speech Research Laboratory for about six months now. We have found the program to be extremely helpful in reducing the amount of time spent in digitizing large numbers of utterances used in creating speech databases. SAP has allowed us to collect a large speech database in a relatively short period of time.

IV. PUBLICATIONS

Papers Published:

Charles-Luce, J. and Dinnsen, D.A. (1987). A reanalysis of Catalan voicing. *Journal of Phonetics, 15*, 187-190.

Connine, C.M. (1987). Constraints on interactive processes in auditory word recognition. *Journal of Memory and Language, 26*, 527-538.

Connine, C.M. and Clifton, C., Jr. (1987). Interactive uses of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 291-299.

Gierut, J.A. (1987). On the assessment of productive phonological knowledge. *National Student Speech, Language, Hearing Association Journal, 14*, 83-100.

Gierut, J.A., and Dinnsen, D.A. (1987). On predicting ease of phonological learning. *Applied Linguistics, 8*, 35-57.

Gierut, J.A., Elbert, M., and Dinnsen, D.A. (1987). A functional analysis of phonological knowledge and generalization learning in misarticulating children. *Journal of Speech and Hearing Research, 30*, 462-479.

Greene, B.G. and Pisoni, D.B. (1988). Perception of synthetic speech by adults and children: Research on processing voice output from text-to-speech systems. In L.E. Bernstein (Ed.), *The Vocally Impaired: Clinical Practice and Research* (pp. 206-248). Philadelphia: Grune & Stratton.

Greenspan, S.L., Nusbaum, H.C., and Pisoni, D.B. (1988). Perceptual learning of synthetic speech produced by rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 421-433.

Gumas, C.C., Hogan, D.L., Oshika, B.L., and Pisoni, D.B. (1987). Evaluation of voice communications systems. *Proceedings of the American Voice Input Output Society*. Sunnyvale, CA: Lockheed.

Luce, P.A. and Pisoni, D.B. (1987). Speech perception: New directions in research, theory, and application. In H. Winitz (Ed.), *Human Communication and Its Disorders* (pp. 1-87). Norwood, NJ: Ablex.

Nusbaum, H. C. and Pisoni, D. B. (1987). Testing the performance of isolated utterance speech recognition devices. *Computer Speech and Language*, *2*, 87-108.

Pisoni, D.B. (1987). Auditory perception of complex sounds: Comparisons of speech vs. nonspeech signals. In W.A. Yost and C.S. Watson (Eds.), *Auditory Processing of Complex Sounds* (pp. 247-256). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pisoni, D.B. (1987). Some measures of intelligibility and comprehension. In J. Allen, D.H. Klatt, and S. Hunnicutt (Eds.), *From Text to Speech: The MITalk System* (pp. 151-171). Cambridge, UK: Cambridge University Press.

Pisoni, D.B. and Luce, P.A. (1987). Acoustic-phonetic representations in the mental lexicon. *Cognition*, *25*, 21-52.

Pisoni, D.B. and Luce, P.A. (1987). Trading relations, acoustic cue integration, and context effects in speech perception. In M.E.H. Schouten (Ed.), *The Psychophysics of Speech Perception* (pp. 155-172). Dordrecht, The Netherlands: Martinus Nijhoff Publishers.

Slowiaczek, L.M., Nusbaum, H.C. and Pisoni, D.B. (1987). Phonological priming in auditory word recognition. *Journal of Experimental Psychology: Human Learning, Memory, and Cognition*, *13*, 64-75.

Slowiaczek, L.M. and Pisoni, D.B. (1987). Speech perception. *McGraw-Hill Encyclopedia of Science and Technology* (4th ed.) (Vol. 17, pp.228-231). NY: McGraw-Hill.

Stemberger, J.P. and Lewis, M. (1987). Reduplication in Ewe. *Phonology Yearbook*, *3*, 151-160.

Summers, W.V. (1987). Effects of stess and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *Journal of the Acoustical Society of America*, *82*, 847-863.

Tomiak, G.R., Mullennix, J.W., and Sawusch, J.R. (1987). Integral processing of phonemes: Evidence for a phonetic mode of perception. *Journal of the Acoustical Society of America*, *81*, 755-764.

Manuscripts Accepted for Publication (In Press):

Connine, C.M., Clifton, C., Jr., and Cutler, A. (in press). Effects of lexical stress on phonetic categorization. *Phonetics.*

Davis, S. (in press). Syllable onsets as a factor in stress rules. *Phonology Yearbook, 5.*

Gierut, J.A. (in press). Maximal opposition approach to phonological treatment. *Journal of Speech and Hearing Disorders.*

Gierut, J.A. and Pisoni, D.B. (in press). Speech perception. In N.Lass, L.McReynolds, J.Northern, and D.Yoder (Ed.), *Handbook of speech-language pathology and audiology.* Philadelphia: B.C. Decker.

Luce, P. A. (in press). Similarity neighborhoods and word frequency effects in visual word identification: Sources of facilitation and inhibition. *Journal of Memory and Language.*

Mullennix, J.W., Greene, B.G., and Pisoni, D.B. (in press). Voice output systems and their perceptual evaluation. In R.J. Porter, Jr. (Ed.) *Speech Applications Handbook.* Hillsdale, NJ: Erlbaum.

Mullennix, J.W. and Pisoni, D.B. (in press). Speech perception: Analysis of biologically significant signals. In R.J. Dooling and S.H. Hulse (Eds.) *The Comparative Psychology of Complex Acoustic Perception.* Hillsdale, NJ: Lawrence Earlbaum Associates.

Pisoni, D.B., Manous, L.M. and Dedina, M.J. (in press). Comprehension of natural and synthetic speech: Effects of predictability on the verification of sentences controlled for intelligibility. *Computer Speech and Language.*

Stemberger, J.P. (in press). Between-word processes in child phonology. *Journal of Child Language.*

Stemberger, J.P. (in press). The reliability and replicability of naturalistic speech error data: A comparison with experimentally induced errors. In B.Baars (Ed.) *The Psychology of Errors: A Window on the Mind?* NY: Plenum.

Summers, W.V. (in press). F1 provides information for final-consonant voicing. *Journal of*

the Acoustical Society of America.

Summers, W.V., Pisoni, D.B., Bernacki, B., Pedlow, R.I., and Stokes, M.A. (in press). Effects of noise on speech production: Acoustic and perceptual analyses. *Journal of the Acoustical Society of America.*

V. Speech Research Laboratory Staff, Associated Faculty, and Technical Personnel
(1/1/87 - 12/31/87)

Research Personnel:

David B. Pisoni, Ph.D. ------------- Professor of Psychology and Director
Beth G. Greene, Ph.D. -------------- Research Scientist and Associate Director

Jan Charles-Luce, Ph.D. ------------ Research Associate
Daniel A. Dinnsen, Ph.D. ----------- Professor of Linguistics
Paul A. Luce, Ph.D. ---------------- Research Associate

Cynthia M. Connine, Ph.D. ---------- NIMH Post-doctoral Fellow*
Stuart A. Davis, Ph.D. ------------- NIH Post-doctoral Fellow
Judith A. Gierut, Ph.D. ------------ NIH Post-doctoral Fellow
John W. Mullennix, Ph.D. ----------- NIH Post-doctoral Fellow
W. Van Summers, Ph.D. -------------- NIH Post-doctoral Fellow

Kazunori Ozawa, B.E.E. ------------- Visiting Scientist†

Michael S. Cluff, B.S. ------------- Graduate Research Assistant
Stephen D. Goldinger, B.A. --------- Graduate Research Assistant
John S. Logan, B.S. ---------------- Graduate Research Assistant
Christopher S. Martin, B.A. -------- Graduate Research Assistant
Robert I. Pedlow, M.Sc. ------------ Graduate Research Assistant
Michael A. Stokes, B.A. ------------ Graduate Research Assistant

Technical Support Personnel:

Cheryl L. Blackerby ---------------- Administrative Secretary
Michael J. Dedina, M.S. ------------ Research Programmer/Assistant
Dennis Feaster, B.A. --------------- Software Development
Jerry C. Forshee, M.A. ------------- Computer Systems Analyst
Luis Hernandez, B.A. --------------- Software Development
David A. Link ---------------------- Electronics Engineer
Gary Link -------------------------- Technical Assistant

Amy Lawlor ------------------------- Undergraduate Research Assistant
Bridget Robinson ------------------- Undergraduate Research Assistant

---

*Now at Department of Psychology, SUNY at Binghamton, Binghamton, NY

†Research Engineer, C&C Information Technology Research Laboratories,
NEC Corporation, Kawasaki, Japan